PROGRAMA DE DOCTORADO EN MATERIALES

TESIS DOCTORAL

Medida de la sección eficaz inclusiva y diferencial de la producción de *quarks top* en asociación con un bosón W en colisiones protón-protón a $\sqrt{s} = 13$ TeV

Measurement of inclusive and differential cross sections for single top quark production in association with a W boson in proton-proton collisions at $\sqrt{s} = 13$ TeV

Víctor Rodríguez Bouza

Universidad de Oviedo

Medida de la sección eficaz inclusiva y diferencial de la producción de *quarks top* en asociación con un bosón W en colisiones protón-protón a $\sqrt{s} = 13$ TeV

Measurement of inclusive and differential cross sections for single top quark production in association with a W boson in proton-proton collisions at $\sqrt{s} = 13$ TeV

Febrero de 2023

Autor

Víctor Rodríguez Bouza

Directores

Bárbara Álvarez González

Isidro González Caballero

Universidad de Oviedo

RESUMEN DEL CONTENIDO DE TESIS DOCTORAL

1.- Título de la Tesis	
Español/Otro Idioma: Medida de la sección eficaz inclusiva y diferencial de la producción de *quarks top* en asociación con un bosón W en colisiones protón-protón a $\sqrt{s}$ = 13 TeV	Inglés: Measurement of inclusive and differential cross sections for single top quark production in association with a W boson in proton-proton collisions at $\sqrt{s}$ = 13 TeV

2.- Autor	
Nombre: Víctor Rodríguez Bouza	
Programa de Doctorado: Programa de Doctorado en Materiales	
Órgano responsable: Centro Internacional de Postgrado	

RESUMEN (en español)

El objetivo principal de esta tesis es el estudio de la producción conjunta de un *quark top* y un bosón W (proceso tW) usando los datos recogidos por el experimento CMS durante el Run-2 del LHC en colisiones protón-protón a una energía en el centro de masas de $\sqrt{s}$ = 13 TeV y correspondientes a una luminosidad total integrada de 138 fb^{-1}.

Se ha determinado la sección eficaz inclusiva de este proceso, obteniendo un valor de 79.2 ± 0.9 (estad.) $^{+7.7}_{-8.0}$ (sist.) ± 1.2 (lumi.) pb. Se trata de medida más precisa hasta este momento del proceso tW con una incertidumbre total de solo el 10%. Esta sección eficaz experimental es compatible con las predicciones del modelo estándar (SM) a aNNLO+NNLL en QCD, y también con las estimaciones recientes a aN^{3}LO+NNLL, también en QCD. Las secciones eficaces diferenciales normalizadas a nivel de partícula para este proceso fueron estimadas como una función de varios observables físicos de los objetos del estado final. Se han comparado los resultados con distintas simulaciones a NLO en QCD, articuladas con dos generadores de cascada de partones diferentes. Existe un acuerdo global con las predicciones del SM. Sin embargo, hay algunas discrepancias que han sido observadas también en otros procesos de producción de *quarks top*. Las pequeñas diferencias entre las predicciones apuntan a un reducido efecto de la interferencia con t$\bar{\text{t}}$ en la región fiducial escogida.

La actualización de fase 2 del detector CMS está motivada por los grandes retos que imponen las duras condiciones previstas cuando empiece a funcionar el LHC de Alta Luminosidad (HL-LHC). En este trabajo se describen los trabajos realizados para caracterizar el envejecimiento esperado en las cámaras de muones, así como el estudio de un nuevo algoritmo de *trigger* que aproveche la nueva electrónica que se instalará.

The main objective of this thesis is the study of the production of a top quark in association with a W boson (tW proccess) using proton-proton collision data collected by the CMS experiment at the LHC during Run 2 at a centre-of-mass energy of $\sqrt{s}$ = 13 TeV and corresponding to a total integrated luminosity of 138 fb^{-1}.

The inclusive cross section of this process was measured obtaining a value of 79.2 ± 0.9 (stat.) $^{+7.7}_{-8.0}$ (syst.) ± 1.2 (lumi.) pb. The total 10% uncertainty achieved provides the best precision so far for this process. This experimental cross section is compatible with the standard model (SM) predictions at aNNLO+NNLL in QCD, and with the recent estimations at aN^{3}LO+NNLL, also in QCD. The normalised differential cross section at particle-level for this process was estimated as a function of several physical observables of the final-state objects. The results are confronted with different simulations at NLO in QCD interfaced with two different parton shower generators. There is an overall agreement with the SM expectations. There are, however, some discrepancies that have been also observed in other top quark production modes. The small differences between the predictions point to small effects due to the interference with $t\bar{t}$ in the fiducial region chosen.

The future Phase-2 upgrade of the CMS detector is driven by the important challenges imposed by the harsher condition expected at the High Luminosity LHC (HL-LHC). A description of the efforts carried out to characterize the ageing of the muon detectors and to study a new trigger algorithm that takes advantage of the new electronics that will be installed is included.

SR. PRESIDENTE DE LA COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO EN MATERIALES

Contents

Contents

List of Figures

List of Tables

List of Acronyms

4FS Four-flavour scheme. 19, 28

5FS Five-flavour scheme. 19

ALICE A Large Ion Collider Experiment. 1, 29

AM Analytical method (trigger algorithm). 60–65, 67, 142, 146, 147

aN^{3}LO Approximately next-to-next-to-next-to-leading order. 24, 111, 139, 143

aNNLO Approximately next-to-next-to-leading order. 24, 111, 139, 143

ASIC Application-specific integrated circuit. 59

ATLAS A Toroidal Large Hadron Collider Apparatus. 1, 19–21, 24, 25, 28–30, 36, 110, 111, 131, 139, 141, 143, 145

AUC Area under the curve. 101, 102

BDT Boosted decision tree. 101, 102, 105, 109, 110

BSM Beyond the standard model. 15, 19, 69

BTI Bunch and track identifier. 60

BX Bunch crossing. 60

CDF Collider Detector at Fermilab. 21, 24

CERN *Conseil européen pour la recherche nucléaire*. 1, 2, 29, 30, 41, 52, 60, 142, 146

CHT Compact Hough transform. 62, 65

CKM Cabibbo-Kobayashi-Maskawa (matrix). 11, 20, 23, 47

CMS Compact Muon Solenoid. 1, 2, 19–21, 24, 25, 29, 30, 32, 34, 36–41, 43, 44, 46, 48–53, 55, 57, 59, 61–65, 67–69, 73, 74, 76, 77, 81–83, 85, 89, 102, 110, 111, 115, 131, 139, 142, 143, 146, 147, 149

CR Colour reconnection. 71, 93

CSC Cathode strip chambers. 40, 44

D0 D-zero experiment. 21, 24

DR Diagram removal (method). 26, 28, 93, 128–131, 142, 146, 149

DS Diagram subtraction (method). 26–28, 93, 128–131, 142, 146, 149

DT Drift tubes. 40, 42, 44, 45, 52, 53, 55, 60, 61, 63–68, 142, 146, 147

DY Drell-Yan (process). 70, 71, 94, 95, 97, 98, 106, 110, 122

EB Electromagnetic barrel calorimeter. 73, 74

ECAL Electromagnetic calorimeter. 36, 38, 39, 45–47, 73–75, 88, 91

EE Electromagnetic endcap calorimeter. 73, 74, 88

EFT Effective field theory. 15

ERD Early resonant decays. 71, 93

EW Electroweak (theory). 4, 10

FPGA Field-programmable gate array. 59, 62, 67

FSR Final-state radiation. 18, 91, 93, 108, 110, 115

GEM Gas electron multiplier. 40

GIF++ Gamma irradiation facility. 142, 146

GOF Goodness of fit (test). 101, 102, 106

GSF Gaussian-sum filter. 45, 46

GUT Grand unified theory. 16

HB Hadronic barrel calorimeter / Hough transform-based (trigger algorithm). 60, 62–65, 74, 142, 146, 147

HCAL Hadronic calorimeter. 36, 39, 45–47, 74, 75, 88

HE Hadronic endcap calorimeter. 74, 88

HEP High-energy physics. 1, 29, 38, 55, 57, 62, 82, 89

HF Heavy-flavour (quark). 82, 91, 92

HL-LHC High-luminosity Large Hadron Collider. 51, 52, 59, 60, 63, 142, 146

HLT High-level trigger. 59, 60, 83

ID Identification (criterium/criteria). 73, 75–78, 81, 84, 85, 90, 91, 94, 115

iRPC Improved resistive plate chambers. 40

ISR Initial-state radiation. 18, 91, 93, 115

JER Jet energy resolution. 92, 111

JES Jet energy scale. 101, 108, 110, 111, 113, 115

KS Kolmogorov-Smirnov (test). 101

L Layer (of drift tube cells). 40

L1 Level one (trigger/jet energy correction step). 59–62, 64, 65, 67, 85, 88, 142, 146

L2 Level two (of jet energy correction step). 85

L3 Level three (of jet energy correction step). 85

LEP Large Electron Positron collider. 29

LHC Large Hadron Collider. 1, 2, 19, 20, 24, 25, 29–34, 36, 38, 40, 41, 51, 55, 59, 61, 74, 82, 84, 89, 139, 141–143, 146

LHCb Large Hadron Collider Beauty. 1, 29–31

LO Leading order. 21, 23, 68, 71

MB Muon barrel (chamber). 63–65, 142, 147

MC Monte Carlo / Monte Carlo (simulation). 2, 25–28, 43, 50, 53, 54, 56, 57, 59, 60, 62, 64, 66, 68–71, 81–86, 88–94, 97–99, 101–106, 108–110, 113, 116, 121–123, 126–128, 141, 142, 145, 146, 149

ME Matrix element. 18, 28, 68, 89, 93, 94, 108, 110, 115, 149

ML Machine learning. 48, 101

MPF Missing projection fraction. 91

MPI Multiple partonic interactions. 18, 70, 71

MVA Multivariate analysis. 101, 106, 116, 131

NLO Next-to-leading order. 24–26, 28, 68, 71, 139, 142, 143, 146, 149

NNLL Next-to-next-to-leading logarithm. 21, 24, 71, 110

NNLO Next-to-next-to-leading order. 21, 23, 71, 110

PDF Parton density function. 17–19, 21, 23–25, 27, 70, 89, 92, 113, 115

PF Particle Flow (algorithm). 43–45, 47, 50, 73, 75, 77, 81, 85

POI Parameter of interest. 100, 105, 121, 124, 125

PS Parton shower. 17, 18, 28, 68, 70, 71, 89, 92–94, 108, 115, 149

PU Pileup. 52, 85, 89–92, 110

PV Primary (interaction) vertex. 73

QCD Quantum chromodynamics. 11–13, 17, 18, 20, 21, 23–26, 71, 85, 93, 111, 139, 142, 143, 146, 149

QED Quantum electrodynamics. 69

QFT Quantum field theory. 3, 11

ROC Receiver operating characteristic (curve). 101, 102

RPC Resistive plate chambers. 40, 41, 44, 53–55, 62–65, 142, 147

SC Supercluster. 73, 75

SF Scale factor. 75, 78, 81, 83–87, 90, 92, 111

SL Superlayer (of drift tube layers). 40, 52, 60–62, 65–67

SM Standard model (of particle physics). 1–7, 10–17, 19, 32, 68, 111, 119, 139, 141, 143, 145

SSB Spontaneous simmetry breaking. 8, 10, 11, 19

SUSY Supersymmetry (theory). 15, 19

T0 Tier zero. 41, 60

T1 Tier one. 41

T2 Tier two. 41

T3 Tier three. 41

TDR Technical design report. 51, 62

TOE Theory of everything. 16

TP Trigger primitive. 60, 62, 64, 65

TRACO Track correlator. 60

TS Trigger server. 61

UE Underlying event. 19, 70, 115

WLCG Worldwide Large Hadron Collider computing grid. 41

WP Working point. 73–77, 81, 86

Introduction

Humankind could be described, at least partially, as a species of problem-solvers. We love mysteries, we take fun in solving puzzles, and (more recently) we also enjoy playing video-games, some of them with a large logical content. And this is understandable, since increasing the comprehension of the surrounding reality gives us tools to live better and, more importantly, to help us understand our own life. To apprehend how reality works helps us understand our role in the Universe as human beings and, in some sense, increase our awareness.

Therefore, it is reasonable that explanations about how reality works have existed for a long time. First, they were shared in the format of tales, myths and legends, and with the birth of philosophy, logical reasoning took over. This allowed the eventual emergence of science (usually considered at the Renaissance, with the appearance of the scientific method). One of the most common questions that humankind asked itself, even when myths were used as a way to share knowledge about reality, was to know the nature of ourselves (and the rest of the universe). Nowadays, the scientific field of particle physics is devoted to that enterprise: to understand the ultimate elements of matter and how they interact among themselves.

The most successful theory that explains this puzzle is the standard model of particle physics (SM), mainly developed in the second half of the 20$^{\text{th}}$ century. We now put it to the test in various laboratories, although one of the most relevant by far is the CERN (the *Conseil européen pour la recherche nucléaire* per its original naming): a high energy physics (HEP) complex near Geneva (Switzerland). It hosts the Large Hadron Collider (LHC), the most powerful particle collider to date. The LHC, since its beginnings, has provided to its experiments, the main being ALICE, ATLAS, CMS, and LHCb, the largest amount of proton-proton collision data ever recorded, and with the highest energy: $\sqrt{s} = 13.6$ TeV in the centre-of-mass frame. In 2012 two of the main detectors at the LHC (CMS and ATLAS) discovered the Higgs boson, the last, long searched, particle of the SM.

The heaviest elementary particle known, the top quark, was discovered in 1995 with the Tevatron proton-antiproton collider at Fermilab (Chicago, USA). This large mass affects its mean lifetime, which is very small, and thus it almost immediately decays. For this reason, we can study its decay products and derive properties from the naked top quark itself, such as its polarisation. This process is experimentally impossible with other quarks created in their final state, as they hadronise to form hadrons and thus the free particle information is convoluted with that of other quarks (effectively loosing it). The large mass indicates also a relevant coupling with the Higgs boson field: that value affects the stability of the Higgs boson propagator, as there are

significant corrections to it that depend on the top quark mass. In addition, there are several beyond-the-SM proposals in which the top quark has a prominent role as a discovery proxy or nexus between new fields/particles and SM ones.

Top quarks may be produced at the LHC through different fundamental processes. The aim of this thesis is to study the production of a single quark top in association with a W boson, also denoted as the tW process. We have measured the inclusive and differential cross-section of the tW process in proton-proton collisions at the LHC at a centre-of-mass energy of $\sqrt{s} = 13$ TeV [1]. We have used data collected by the CMS detector at CERN in 2016, 2017 and 2018, amounting to an integrated luminosity of 138 fb^{-1}. The differential analysis described in this thesis had also a preliminary result with 2016-only data [2] (roughly $\frac{1}{4}$ of the final result's integrated luminosity). With that same data set, the inclusive cross section was also measured [3]: the main motivation for the extension of the measurement at the same energy lies on the updated knowledge of the detector, as well as uncertainty estimations.

This document is structured as follows. Chapter 1 gives an introduction to the theoretical context needed, such as the SM, proton-proton collision phenomenology, or top quark physics. The description of the experimental setup, with the LHC and CMS is in Chapter 2. The contributions to the maintenance and operation of the detector are also detailed there. Most of the remaining chapters are devoted to the tW process measurement: the trigger selection, data, and Monte Carlo (MC) simulations used are described in Chapter 3. Chapter 4 contains the object identification and corrections, whereas Chapter 5 explain the different sources of uncertainty considered in this study. The details on the inclusive cross section measurement may be found in Chapter 6, together with a comprehensive description of the event selection. Afterwards, the differential one is treated in Chapter 7. Finally, the summary and conclusions are outlined in Chapter 8, and their translation to Spanish in Chapter 9. The Appendix A shows introductory studies on the $b\bar{b}\ell^+\nu_{\ell+}\ell^-\nu_{\ell-}$ sample of the WWbb process, that agglutinates both tW and the top-antitop quark pair production, $t\bar{t}$.

1 Overview of collider particle physics

I N this chapter we will briefly discuss the current theoretical framework used in high energy physics: the SM. A short review of this theory's limitations and its potential expansions follows, before a description of how to experimentally understand particle collisions is given. The last section is devoted to a brief overview of top quark physics in hadron-hadron colliders.

1.1 The standard model of particle physics

The SM is a quantum field theory (QFT) that explains how the ultimate components of matter interact among themselves through three of the four elemental forces of the universe: the electromagnetic, weak nuclear, and strong nuclear interactions. This theory has proven to be very successful, experimentally speaking, with continuous confirmations of their predictions in the past decades.

Being a QFT, it can incorporate two of the new perspectives that the beginning of the 20$^{\text{th}}$ century brought. First, a (special) relativistic framework, that allows us to work with very high energies, and secondly, a quantum environment: essential to properly understand the small scales where we want to work. When considering a (special) relativistic quantum mechanics environment, antiparticles are essential to avoid breaking causality. And finally, as the number of particles cannot be correctly fixed, due to Heisenberg's uncertainty principle, the concept of **field** arises as a necessity. This also allows us to better comprehend interactions between particles, as inelastic scattering or decays could hardly be understood with classical mechanical "marble-based" layouts. Considering particles as excitations of fields solve these issues.

The fields described by the SM are either fermionic (with non-integer spin and following the Fermi-Dirac statistics) or bosonic (with integer spin and following the Bose-Einstein statistics). Fermions end up making matter as we know it (baryonic matter, to be precise), and are divided in leptons and quarks. The division is caused by the different forces through which they can interact with other fields: the particles can interact using a specific force if they have the associated charge to that interaction.

In the case of the quarks, they can interact through the strong, weak and electromagnetic forces, whereas leptons cannot feel the strong interaction. Quarks come in pairs: one with a positive electric charge of $\frac{2}{3}e$, being e the amount carried positively by a proton (or negatively by an electron) and other with a negative one of $-\frac{1}{3}e$. They also have colour charge (the one

from the strong force). Leptons come also in pairs: one with an electric charge of $-e$ and other, called neutrino that is neutral and thus does not interact through the electromagnetic charge. Neutrinos, as well as all leptons, do carry flavour charge and consequently they can interact through the weak force. There are three "copies" of these pairs of quarks and leptons called generations: their properties of all these fields are identical, except for their masses. As the SM respects Lorentz transformations, all fields are representations of the Lorentz group: in the case of fermions, they are Dirac' spinors (a mixture of a couple of Weyl spinors: one lefthanded and the other righthanded).

The boson fields can be classified depending on their spin. The Higgs boson, without electric charge, is the only one that is scalar (with zero spin), and, through the interaction with it, fermions obtain its masses (that are a measure of how the interaction with this field is). The vector bosons (all of them with spin 1) are said to mediate the fundamental forces: gluons are massless and do not possess electric charge, but they have colour charge and allow interactions through the strong force. Photons are also massless, and neutrals, and they carry the electromagnetic force. The remaining bosons, the W$^+$, W$^-$, and the Z bosons mediate the weak interaction: they have mass, and the W$^\pm$ bosons are electrically charged.

The SM interactions are defined through a Lagrangian density $\mathcal{L}(\psi_i, \partial_\mu \psi_i, x^\mu)$, where ψ_i are the fields, and x^μ the spacetime coordinates. In order to allow creating predictions that we can contrast with data, we shall look for a renormalisable Lagrangian density: this constrains the terms that can appear in it. In particular, all the terms will be of, at maximum, order four. The terms present in this Lagrangian come from the different "sectors", as usually called, that exist: these parts of the SM depend on the interactions they describe. Most of the Universe's symmetries are global (such as the baryonic or leptonic numbers), but some of them are promoted to local, as we will see in the following subsections. The field contents of the SM are shown in Fig. 1.1, where antiparticle fields are omitted and the different gluons (due to their colour charge content) are shown as one.

1.1.1 Electroweak sector

The electric and nuclear weak forces are contained inside the electroweak (EW) sector. Both interactions are grouped together in a comprehensive theory that was developed by Glashow, Weinberg, and Salam[5, 6, 7]. The model derives from a local gauge symmetry based on interactions derived from a $SU(2)_L \times U(1)_Y$ Lie group, where $SU(2)$ refers to the special unitary group of second degree, and $U(1)$ to the unitary group of first degree. The "L" refers to *left* (sometimes "W" appears there, from *weak*) and the "Y" to the hypercharge that we will describe later.

As we mentioned before, Dirac's fermions are made of two Weyl spinors, one of them left-handed and the other right-handed ($\psi = \psi_L + \psi_R$, $\psi_L = \frac{1-\gamma_5}{2}\psi$, $\psi_R = \frac{1+\gamma_5}{2}\psi$)[1]. As theoretically we could build a model like this one to an arbitrary number of generations, we will focus, with no

---

[1] γ_5 is the product of the Dirac matrices, $\gamma_5 := \gamma_0 \cdot \gamma_1 \cdot \gamma_2 \cdot \gamma_3$.

Standard Model of Elementary Particles
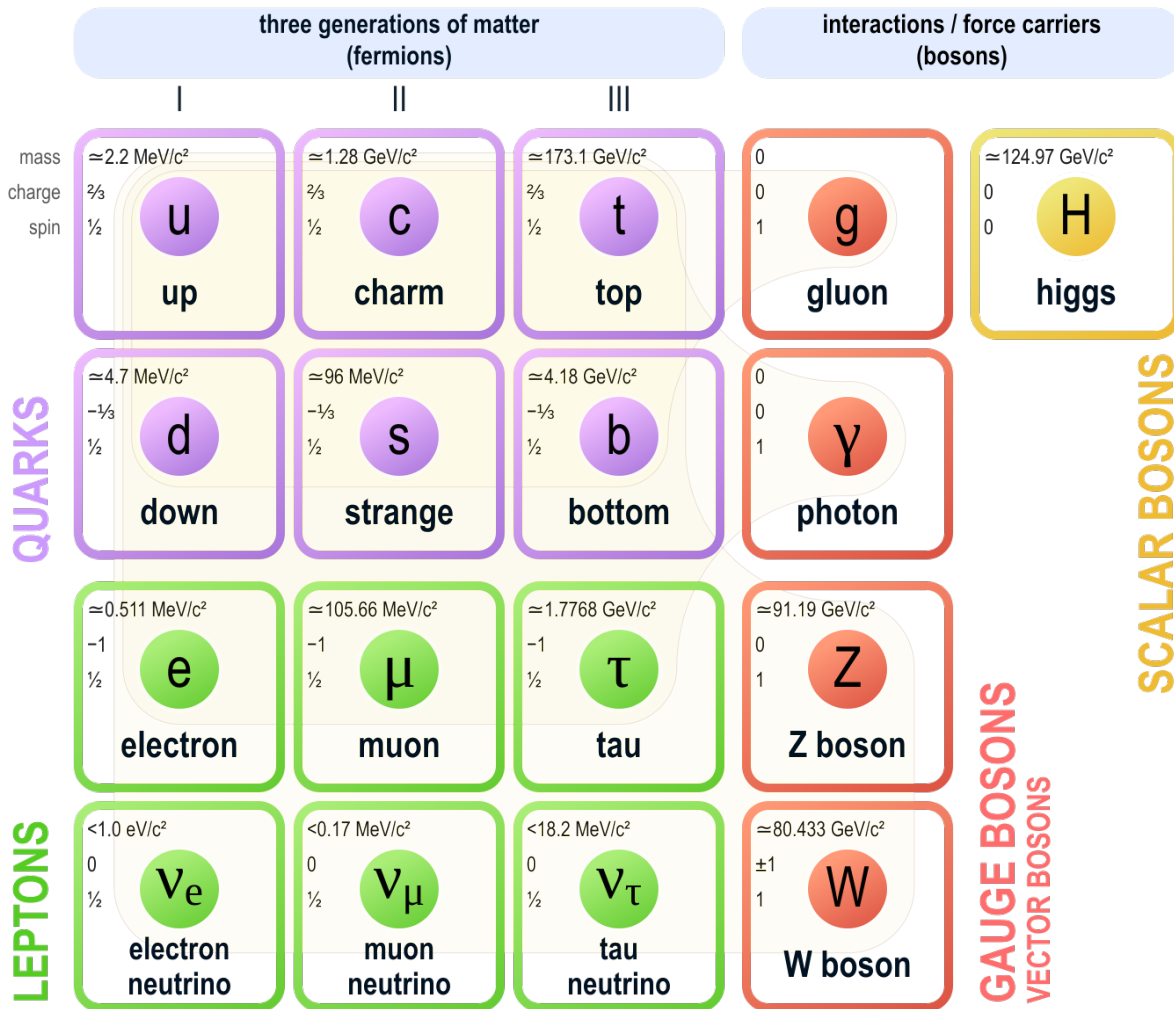
Figure 1.1: [4] Elementary particles of the SM of particle physics.

loss of generality, on the first generation of fermions only. Starting with the leptons, we will work with generic e and ν. The $SU(2)$ part of the theory, that aims to explain the weak interaction, is set to act only over the left-handed fermions, and at the same time over both (e and ν). This reflects the consequences of the Wu experiment [8], that proved that the weak interaction did not respect parity and thus any theory that might attempt to explain it must be chiral by construction. To model this, we construct doublets of fermions that are left-handed, whereas we leave the right-handed elements alone (as singlets):

$$L = \begin{pmatrix} e_L \\ \nu_L \end{pmatrix}, \quad e_R. \tag{1.1}$$

Right-handed neutrinos do not interact at all inside the SM and therefore are usually not incorporated. However, after the confirmation that neutrinos do have mass, mixing between left-handed and right-handed components can occur.

The $SU(2)$ symmetry needs of the existence of three gauge bosons so that we can define the actual gauge transformation under which the Lagrangian is invariant. These are three vector fields: W_μ^i with $i = 1, 2, 3$ and with the μ index going over spacetime components. The necessary covariant derivative is constructed by using their strength tensor,

$$W_{\mu\nu}^i = \partial_\mu W_\nu^i - \partial_\nu W_\mu^i + g_W \epsilon_{ijk} W_\mu^j W_\nu^k, \tag{1.2}$$

where g is the coupling constant associated with the transformation related to $SU(2)$. With the help of the Pauli matrices, generators of $SU(2)$, we obtain the covariant derivative:

$$D_\mu = \partial_\mu - ig T_3 \sigma_i W_\mu^i. \tag{1.3}$$

With these ingredients, the Lagrangian density can be proven at this point to be the following (with $\slashed{D} = \gamma^\mu D_\mu$),

$$\mathcal{L} = -\frac{1}{4} W_{\mu\nu}^i W_i^{\mu\nu} + i \overline{L} \slashed{D} L. \tag{1.4}$$

Due to Noether's theorem, a conserved quantity must exist because of these local gauge symmetries. The value, or charge, associated with the $SU(2)$ is called weak isospin, T_3. Its name makes reference to the chiral nature of the $SU(2)$ interaction we just built. It is related with the charge of the $U(1)$ group, called the hypercharge (Y), by the Gell-Mann–Nishijima relation [9, 10], that clarifies which values of weak isospin and hypercharge are allowed, depending on the electric charge values (Q):

$$Q = T_3 + Y. \tag{1.5}$$

A summary of the electric, weak hypercharge, and weak isospin charges can be seen in Tab. 1.1.

Left-chiral fermion fields	Electric charge	Weak isospin	Weak hypercharge
ν_e, ν_μ, ν_τ	0	$+\frac{1}{2}$	-1
e, μ, τ	-1	$-\frac{1}{2}$	-1
u, c, t	$+\frac{2}{3}$	$+\frac{1}{2}$	$+\frac{1}{3}$
d, s, b	$-\frac{1}{3}$	$-\frac{1}{2}$	$+\frac{1}{3}$

Table 1.1: Electric, weak hypercharge, and weak isospin charges values for all left-chiral fermion fields of the SM.

An analogous development can be done for the $U(1)_Y$ part that yields in this case only one vector field B_μ, with an associated coupling constant g'. In this case, the covariant derivative, gauge field strength tensor and Lagrangian density are:

$$
\begin{aligned}
D_\mu =& \partial_\mu - ig'YB_\mu, \\
B_{\mu\nu} =& \partial_\mu B^\nu - \partial_\nu B_\mu, \\
\mathcal{L} =& -\frac{1}{4}B_{\mu\nu}^i B_i^{\mu\nu} + i\overline{L}\slashed{D}L + i\overline{e_R}\slashed{D}e_R.
\end{aligned}
\tag{1.6}
$$

And, the entire electroweak Lagrangian density can be re-written with a combined covariant derivative to this:

$$
\begin{aligned}
D_\mu =& \partial_\mu - igT_3\overrightarrow{T}\cdot\overrightarrow{W_\mu} - ig'YB_\mu, \\
\mathcal{L}_{\text{EWK}} =& -\frac{1}{4}W_{\mu\nu}^i W_i^{\mu\nu} - \frac{1}{4}B_{\mu\nu}^i B_i^{\mu\nu} + i\overline{L}\slashed{D}L + i\overline{e_R}\slashed{D}e_R.
\end{aligned}
\tag{1.7}
$$

We should note at this point that any mass term either for the fermions ($m\overline{\psi}\psi$) or for any of the four gauge bosons are forbidden as they are not invariant under the corresponding gauge symmetries: thus all fermions and also gauge bosons are massless. As we know that these particles have mass, we need a way to provide it in the model.

1.1.1.1 Spontaneous electroweak symmetry breaking

The SM explains mass of bosons and fermions through the commonly called Brout-Englert-Higgs mechanism[2] [12, 13, 14]. This method requires of the spontaneous "breaking" of the local gauge symmetry of $SU(2)_L \times U(1)_Y$ (i.e. the electroweak symmetry) we just described when including a scalar field doublet Φ, defined as

$$
\Phi := \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix},
\tag{1.8}
$$

where ϕ^+ and ϕ^0 are the two components of it: one with positive electric charge, and the other

---

[2]Given that this method was independently published almost simultaneously by multiple groups, it is called of multiple ways: the Brout-Englert-Higgs mechanism, the Higgs mechanism, or even the ABEGHHK'tH mechanism by Higgs himself [11] (for Anderson, Brout, Englert, Guralnik, Hagen, Higgs, Kibble and t'Hooft).

neutral, fixing the hypercharge of the doublet to $+\frac{1}{2}$. This doublet is given an associated Lagrangian density with a kinematic and potential part, as well as an interaction part with the fermions. With the same covariant derivative as before (the one we had in $SU(2)_L \times U(1)_Y$), we can write these extra terms as

$$\mathcal{L}_{\text{scalar}} = (D\mu\Phi)^\dagger(D_\mu\Phi) - V(\Phi^\dagger\Phi) - G \cdot (\overline{L}\Phi e_R + \overline{e_R}\Phi^\dagger L), \qquad (1.9)$$

being all of them gauge invariant and renormalisable. It should be noted that the scalar field can couple to both the left-handed doublets and right-handed singlets, being the G its coupling constant (a number we can always define as real, by absorption of any complex phase in the definitions of L, Φ or e_R). The potential $V(\Phi^\dagger\Phi)$ is defined as

$$V(\Phi^\dagger\Phi) = \mu^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2, \qquad (1.10)$$

with $\mu^2 < 0$ and $\lambda > 0$. The shape of this potential is the known hat-like shape or *sombrero* potential, as shown in Figure 1.2. The symmetric solution, centred, is clearly unstable and therefore, *spontaneously*, it can fall to one of the (many) stable solutions, fixing one of them as the physical one. For this case, with an abelian local gauge symmetry like $SU(2)_L \times U(1)_Y$, our symmetry is broken (no generators from $SU(2)_L$ nor $U(1)_Y$ respect the vacuum). Through the Higgs mechanism, we can fix our gauge such that this breaking also yields new combinations of the original $\overrightarrow{W_\mu}$ and B_μ with the components of the scalar field: three massive and one that is not. In addition, one degree of freedom from the previous scalar field remains that acquires mass. The Lagrangian density, with the corresponding change of variable already done ends up being

$$\mathcal{L}_{\text{EWSM}} = -\frac{1}{4}W^i_{\mu\nu}W^{\mu\nu}_i - \frac{1}{4}B^i_{\mu\nu}B^{\mu\nu}_i + i\overline{L}\slashed{D}L + i\overline{e_R}\slashed{D}e_R +$$
$$\left[D\mu\begin{pmatrix}0\\\frac{v+H}{\sqrt{2}}\end{pmatrix}\right]^\dagger\left[D_\mu\begin{pmatrix}0\\\frac{v+H}{\sqrt{2}}\end{pmatrix}\right] - V\left(\frac{v+H}{\sqrt{2}}\right) - G \cdot \left[\overline{L}\begin{pmatrix}0\\\frac{v+H}{\sqrt{2}}\end{pmatrix}e_R + \overline{e_R}\begin{pmatrix}0\\\frac{v+H}{\sqrt{2}}\end{pmatrix}^\dagger L\right],$$
$$(1.11)$$

with the scalar interaction term usually called the Yukawa Lagrangian, v the vacuum expectation value and H the new scalar field that acquires mass. The scalar kinetic terms contain hidden the mix details of the original scalar field (before the spontaneous symmetry breaking or SSB) with the $\overrightarrow{W_\mu}$ and B_μ bosons. It can be seen that, when expanding them, one can get[3]:

$$\frac{v^2}{8} \cdot \left(W^{(1)}_\mu, W^{(2)}_\mu, W^{(3)}_\mu, B_\mu\right) \cdot \begin{pmatrix}g^2 & 0 & 0 & 0\\0 & g^2 & 0 & 0\\0 & 0 & g^2 & -gg'\\0 & 0 & -gg' & g'^2\end{pmatrix}. \qquad (1.12)$$

---

[3] We note with $W^{(1)}_\mu$ the first element of the $\overrightarrow{W_\mu}$ vector, and so on: they are not powers of them.
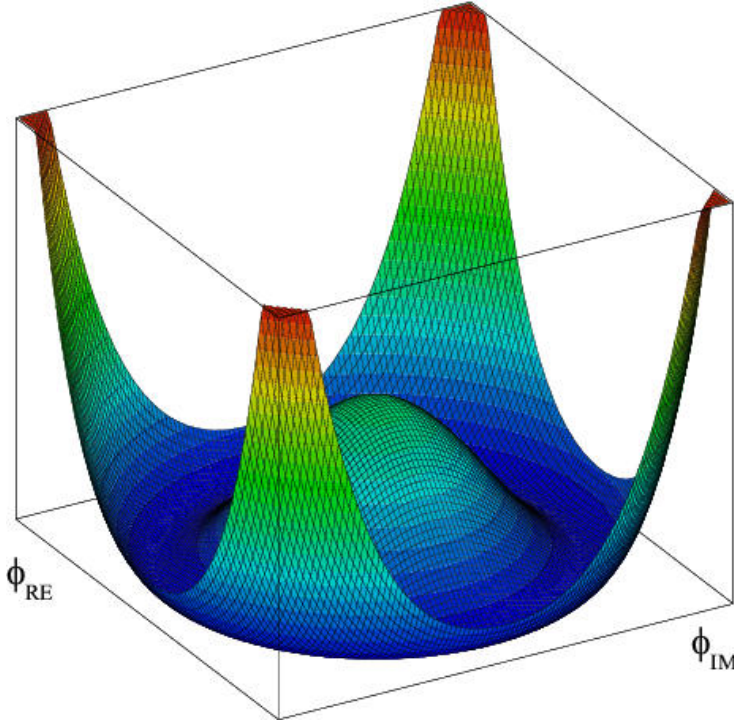
Figure 1.2: Graph showing the so-called *sombrero* potential, from Eq. (1.10).

The third element of $\overrightarrow{W}_\mu$ and B_μ are mixed: they can be disentangled by diagonalising the matrix. Being the submatrix that mixes them, a symmetric 2×2 one, its diagonalisation is just a rotation and therefore we can express that "turn" with an angle. The eigenvalues end up being zero for the combination noted by A_μ and $\frac{v^2}{8}(g^2 + g'^2)$ by the one commonly called Z_μ. These combinations are described by the weak mixing or Weinberg angle θ_W:

$$
\begin{aligned}
A_\mu &= \sin\theta_W W_\mu^{(3)} + \cos\theta_W B_\mu, \\
Z_\mu &= \cos\theta_W W_\mu^{(3)} + \sin\theta_W B_\mu, \\
\tan\theta_W &= \frac{g'}{g}.
\end{aligned}
\tag{1.13}
$$

The A_μ field is the photon (γ) field, and the Z_μ the also observed Z^0 boson, and they are neutral (if not, the vacuum would be charged). From its eigenvalue, the mass of the Z boson ends up being $m_Z = \frac{v}{2}\sqrt{g^2 + g'^2}$. The $W^\pm$ bosons can be obtained from the mixing of $W^{(1)}$ and $W^{(2)}$, as

$$
W_\mu^\pm := \frac{1}{\sqrt{2}}\left(W_\mu^{(1)} \mp iW_\mu^{(2)}\right),
\tag{1.14}
$$

with the same mass that from Eq. (1.12) is $m_W = \frac{1}{2}gv$. From these relations, one can state that the Z boson is heavier than the W. Even more,

$$m_W = m_Z \cdot \cos\theta_W. \tag{1.15}$$

Thus, the boson ($W^\pm$, Z^0, and γ) masses are fixed by the SSB of the EW symmetry, and depend on the vacuum expectation value v and the coupling constants g and g'. This does not apply to fermions, and to get their mass terms we need to review Yukawa's Lagrangian. For leptons, we can obtain terms like the following:

$$-G\frac{v}{\sqrt{2}}\left(\overline{e_L}e_R + \overline{e_R}e_L\right) = -G\frac{v}{\sqrt{2}}\bar{e}e = -m_e\bar{e}e \Rightarrow m_e = G\frac{v}{\sqrt{2}}. \tag{1.16}$$

Consequently the electron mass also depends on v, but also on the coupling constants G, commonly called Yukawa coupling constants: a way to measure the strength of the Higgs boson field interactions with other fields. Neutrinos do not have mass in the SM (because they do not have a corresponding right-handed neutrino to which they could couple).

Finally, the mass from the H scalar field, that will end up being the Higgs boson field, can be extracted from the potential part, obtaining a term like

$$+\mu^2 H^2 =: -\frac{1}{2}m_H^2 H^2 \Rightarrow m_H = \sqrt{-2\mu^2}. \tag{1.17}$$

Thus, interestingly, the Higgs' boson mass does not depend on the vacuum expectation value: only on μ.

1.1.1.2 Adding quarks and other generations: the CKM matrix

Until now, we have considered only two leptons (the electron and its neutrino), given that the generalisation from them to quarks and other generations is straight-forward. Quarks can be considered, once again, as $SU(2)_L$ doublets and singlets, e.g. considering the first generation of quarks (u and d)

$$\begin{pmatrix} u_L \\ d_L \end{pmatrix}, u_R, d_R. \tag{1.18}$$

This can be done also for the second generation (μ, ν_μ, c, s) and the third one (τ, ν_τ, t, b). There is however a subtlety when adding them to the Lagrangian density: now, the Yukawa coupling G are not mere numbers, but a matrix, and as we had to do for the $W_\mu^{(3)}$ and B_μ, the mass eigenstates might not be the same as the ones of the Yukawa interaction. For the case of the leptons as we described them, this is not an issue (as neutrinos do not have mass in the SM), and we have only three Yukawa couplings to the electron, muon, and tau leptons. Notwithstanding, for quarks this does not happen: the common approach is to diagonalise the mass matrix, obtaining clear mass terms in the Lagrangian, at the price of having a weak interaction that mixes the generations between them. This mixing is parameterised by the Cabibbo-Kobayashi-Maskawa (CKM) matrix,

$$\begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \tag{1.19}$$

1.1.2 Strong sector

The existence of a new force, different from gravity, electromagnetic and weak, was originally hypothesised in the second half of the 20th century to explain how, despite the electromagnetic repulsion among protons, nuclei are bonded together. This interaction, necessarily *stronger* than the electromagnetic (at least in the nucleus' context) was explained in a QFT developed by Fritzsch, Leutwyler and Gell-Mann [15], based on the previous work of Yang and Mills [16].

This strong interaction, that among fermions, only quarks feel, is explained in this theory with the a new local gauge symmetry, constructed with the Lie group $SU(3)_C$. The "C" refers to colour[4], which is the name of the associated charge to the strong force, also called colour force. The charge also gives name to the theory, called quantum chromodynamics (mirroring quantum electrodynamics) or QCD.

A development relatively similar to what was explained with the electroweak sector of the SM can be done here, obtaining a QCD Lagrangian density from the strength tensor field $G_{\mu\nu}^a$ as follows.

$$\mathcal{L}_{\text{QCD}} = -\frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu}, \quad G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_s f_{bc}^a G_\mu^b G_\nu^c \tag{1.20}$$

Now, with the $SU(3)_C$ group we have eight generators (the Gell-Mann matrices T^a; all Arabic indexes go $1, 2, \ldots, 8$) that give rise to our eight vector bosons G_μ^a, called gluons. As they do not interact with the Higgs field, they do not get mass from the SSB of the electroweak symmetry and are massless. g_s is the coupling constant of the strong force and f^{abc} are the structure constants of the $SU(3)$ group. The covariant derivative of the SM, from Eq. (1.7), is updated to include QCD with an extra term $-ig_s C T^a G_\mu^a$, where C is the colour charge. The addition of the QCD Lagrangian density from Eq. (1.20) to the electroweak one from Eq. 1.7 and the mentioned update of the covariant derivative finishes the construction and description of the SM of particle physics.

The fact that this theory is constructed around a non-abelian symmetry affects it qualitatively, providing the last term that appears in the Lagrangian from Eq. 1.20. Thanks to it, self-interactions between gluons are not forbidden like with the weak or electromagnetic forces (allowing three-particle or four-particle interaction vertexes).

---

[4]It should be noted, as it is done in any text of QFTs or in any SM description, that this charge has not any relationship with the concept of "colour" that we usually use.

1.1.3 Coupling constants and differences among forces

Due to the quantum nature of the Universe encoded in the SM, the concept of virtual particles (or "off-shell" particles) is not a hard-magic element from a complex Brandon Sanderson's fantasy saga [17], but a necessity to correctly explain many physical processes. These excitations of the fields of the SM, that do not respect $p^2 = E^2 - m^2$, can exist under the protection of Heisenberg's uncertainty principle. In particular, when one considers processes that occur at higher and higher energies, corrections done by virtual particles affect more the interactions and should be taken into account to give a correct description (and predictions) of the process.

A way of doing this coherently for different scales of energy is to absorb those variations into a new coupling (no longer constant), called **effective coupling**. Knowing how coupling constants "run" (change) with the energy is essential to give predictions at different scales.

As we saw in the previous section, the three interactions that are part of the SM are associated to different local gauge symmetries, that explain their behaviour in the theory. The weak and electromagnetic forces are related to a $SU(2)_L \times U(1)_Y$ symmetry (that has been broken) whereas a $SU(3)_C$ explains the nuclear strong interaction. The differences between these symmetries reflect the very differences among forces, but also the running of their couplings. This running is usually described through **beta functions**, defined as

$$\beta(g) := \frac{\partial g}{\partial \log \mu},$$

(1.21)

where g is an (effective) coupling and μ is the scale of energy. For the case of quantum electrodynamics, and considering one-loop corrections, this function is (in terms of the fine structure constant α):

$$\beta(\alpha) := \frac{2\alpha^2}{3\pi}.$$

(1.22)

The one for the weak coupling can be extracted from e. The corresponding to the strong force also for one-loop is:

$$\beta(\alpha_s) \approx -5.7 \frac{\alpha_s^2}{2\pi}.$$

(1.23)

Here we can see another relevant difference derived from the nature of the different forces: whereas the electromagnetic and weak beta functions are positive (indicating that the effective coupling increases with energy), for the strong force this is the opposite.

This explains another particularity of the latter: the phenomena of colour confinement and asymptotic freedom. The first occurs when treating QCD processes at low energy: the coupling is so strong that the theory is no longer perturbative and also induces the spontaneous creation of quark-antiquark pairs, effectively preventing freely propagating gluons or quarks to remain that way. A similar thing occurs with the electromagnetic and weak interactions, but with very high
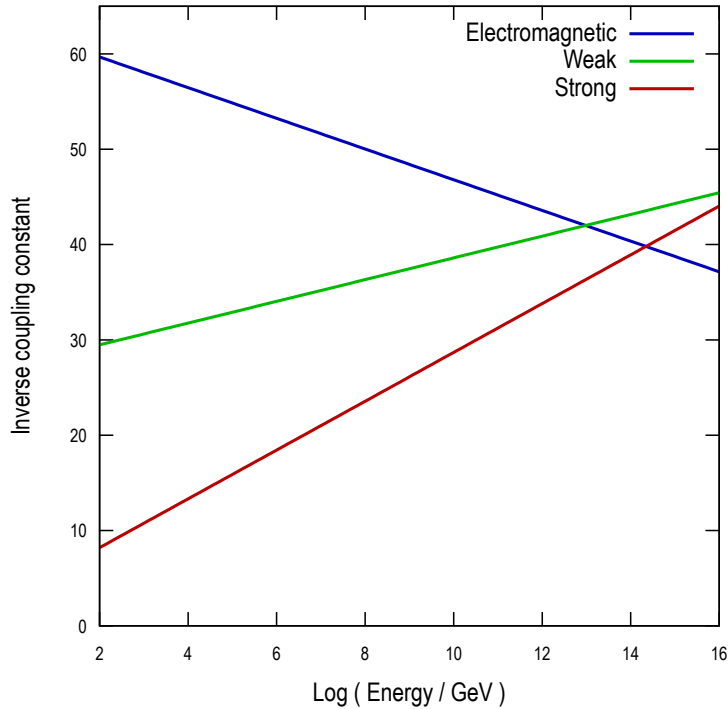
Figure 1.3: [18] Evolution of the inverse of the coupling constants of the electromagnetic, weak, and strong fundamental forces with respect to the logarithm of energy.

energies: they reach a divergence in the coupling (a Landau pole) where the theory is no longer perturbative. The second phenomenon, asymptotic freedom, is the opposite: in high energy values, QCD is perturbative. An evolution of the coupling constants with the energy of the electromagnetic, weak, and strong forces can be seen in Fig. 1.3.

1.2 Frontiers and beyond

Despite being one of the most successful theories in the history of physics, with decades of experimental validations of its predictions, there are aspects from nature that the SM does not provide an answer for, in addition to some experimental results that do not agree with its expectations. Proposals have been made in the last decades to explain these phenomena.

1.2.1 The standard model shortages

The following is a summarised list of the most relevant features with which the SM cannot cope, or was not designed to comply.

Neutrino masses It has been proven that neutrinos change flavour (oscillate) [19, 20] and therefore they have mass. However SM neutrinos are massless by construction, as they do not couple to a right-handed neutrino.

Hierarchy When considering perturbative corrections to the propagator of the Higgs boson, one finds that the size of these corrections is vastly large compared with the scale of the mass of the boson itself, of the order of the Planck mass ($\approx 10^{19}$ GeV). This implies that to get the experimental results we observe, in the framework of the SM one must tweak parameters to a high level of adjustment, so that the large corrections are cancelled. This extreme fine-tuning seems unexpected, and also a possible artifact of the current SM that might not reflect the reality of Nature.

Incompleteness: gravity As we stated at the beginning of Section 1.1, the SM explains interactions between elementary particles depending on *three* of the four fundamental forces of the Universe: the electromagnetic, and the weak and strong nuclear interactions. The lack of gravity in this framework has always been a clear shortage of the SM.

Incompleteness: dark matter Since the first half of the 20$^{\text{th}}$ century, there have been experimental observations according to which the apparent "common" or baryonic matter cannot explain the whole amount of mass of the Universe. Even in the late 19$^{\text{th}}$ century, Lord Kelvin suggested that the observable (visible) mass of stars does not account for all the mass of the galaxy [21]. The work done by Rubin, Ford and Freeman in the 60s and 70s [22, 23] provided strong evidence to the scientific community of the unknown origin of a large portion of the mass of the galaxy, called "dark matter" (as apparently it does not interact with the electromagnetic force). We know now that, from the entire mass-energy content of the Universe, while baryonic matter accounts for $\approx 5\%$, dark matter provides the $\approx 25\%$. While the current SM provides candidates to explain the baryonic content, it does not have any suitable particle to explain the dark matter one.

Incompleteness: dark energy In the late 90s, results from supernovae observations [24, 25] pointed to the fact that the universe is accelerating. These measurements, and others of different nature (e.g. cosmic microwave background) are consistent with the standard model of cosmology: the Λ_{CDM}[5] model, and suggest that the remaining $\approx 7\%$ of the mass-energy content of the Universe corresponds to "dark energy". This contribution, represented in the Λ_{CDM} model as Λ (the cosmological constant), is of unknown nature and it is the responsible for the acceleration of the Universe. The SM does have a candidate to explain this: the vacuum energy density. However, the predicted values from the SM yield a difference of 120 orders (10^{120}) of magnitude with respect to the observed number.

Baryogenesis We know both matter and antimatter exist, but clearly we live in a matter-dominated Universe, with a negligible amount of antimatter on it. Any attempt to create a model of the ultimate elements of matter should explain how we can reach this unbalance, through the baryogenesis (the process that would have created this asymmetry). The SM

---

[5]Lambda cold dark matter.

can accommodate baryogenesis, though it is not proven that the amount of asymmetry that provides is sufficient to explain the current Universe.

1.2.2 Beyond the standard model

Many new theories and models have been proposed to address the mentioned lacks of the SM, as well as others we did not mention. These suggestions that expand the framework are usually called "beyond the standard model" theories, or BSM.

For example, to incorporate neutrinos to the SM, there are alternatives such as seesaw mechanisms [26] that allow to add them along with right-handed neutrinos. However, probably the most known BSM theory that also could in principle be discovered with the current experiments is supersymmetry (SUSY), first envisioned by Miyazawa in 1966 [27]. SUSY theories extend the SM adding an extra symmetry that transforms fermions into bosons. With it, one can expect supersymmetric particles: counterparts to current SM particles, but with different spin (and other properties, such as masses). When the SM is extended with SUSY, issues such as the hierarchy problem can be solved. It also can provide candidates for dark matter (such as the neutralino), and offer a convergence of the gauge coupling constants of the three interactions inside the SM.

There are many hypothetical particles (e.g. axions or sphalerons) and models that conform the BSM category apart from the ones we mentioned. Inside this very diverse jungle of ideas, there are some that add an extra particle or field to the SM, or that also try to be a model that *just* works in a particular range of energies, such as the effective field theories (EFT).

However, and slightly with the same spirit as we began this chapter, there is a clear pathway where many of the theories lie, and attempt to walk: an effort to get explanations that can create a comprehensive description of Nature by incorporating some (if not all) of the current unknowns and shortages inside them. This ambitious enterprise of Humankind is old, even ancient. As a "recent" example in 1814, and obviously within its own context, Laplace envisioned what an entity with enough intellect could do when combining mechanics and gravitation [28]:

> *An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes.*

Theories that attempt to explain as only one force the SM interactions (electromagnetic, weak and strong) are called grand unified theories, or GUTs. Examples of these are those based e.g. in the $SU(5)$ Lie group. However, the most ambitious aim is to obtain the theory of everything (TOE), that would explain not only the interactions and particles a GUT would do, but also gravitation in one unique and coherent framework. Such a theory would include thus the physics developments

from the 20th century that separately helped us fundamentally change our knowledge of reality: quantum physics and general relativity. One of the most known candidates to be a TOE is string theory.

1.3 Phenomenology at proton-proton colliders

This section is intended to serve as a bridge between the theoretical prolegomena we have just discussed and the experimental work that is shown later in the thesis. This gap is not crossed easily, as there are many aspects to consider when doing actual predictions that can be compared with data. We shall give an overview of the necessary topics.

1.3.1 Understanding collisions

The predictions we use in high energy physics for collider experiments are based in the scattering matrix (S-matrix) formalism. This development, from the first half of the 20th century, allows to explain interactions between a set of free incoming particles on their mass-shell that yield another set (not necessarily the same) of free outgoing particles. This S-matrix links all possible interactions between them, even the non-interaction case. The possible combinations due to virtual particles make the total potential interactions infinite. Therefore, pragmatically, a perturbative approach is done depending on the number of interaction vertexes (and thus on the couplings of the forces). The dependence on energy (running) of the couplings is not transferred to the cross section, because a renormalisation can be done when all (infinite) orders are taken into account. However, these perturbative series are done up to a particular (or "fixed") order for each calculation, as it is not easy to incorporate and calculate an arbitrary large number of orders. Thus, to obtain a definite prediction, we must provide a renormalisation scale, μ_R, to the calculation.

One of the most relevant physical observables (if not the most important of all of them) in high energy physics is the **cross section**: a measure of the probability that one particular process takes place. It is given in surface units called barns: this is usually interpreted also as the area where two particles should collide for such an interaction to take place. It depends on the conditions of the particles that interact (usually redefined as a function of the centre-of-mass energy of the collision), but it is independent of the rest of the experimental set-up. It provides the theory expectation, and thus by measuring it, we can confront predictions with reality.

As protons are not elemental particles, we calculate cross sections between their components. The information of the momenta of the proton components or partons is usually encoded in parton density functions (PDFs), that have been experimentally obtained from deep inelastic measurements. The PDFs describe the probability to see a parton with a precise momentum in an interaction given at a particular energy. They are very hard to calculate directly from the SM formalism, as non-perturbative QCD calculations are needed. Thus, to be able to do predictions in proton-proton collisions, a separation at a particular energy called the factorisation

scale, μ_F, is done between the partons coming from the protons (the "soft" process) and the rest of the interaction (the "hard" process). This division allows us to use PDF measured in other experiments with our predictions and thus calculate only the necessary information for the hard process. A representation of a proton-proton collision, with both hard and soft processes, is shown in Fig. 1.4, including the interactions that are described in the following paragraphs.
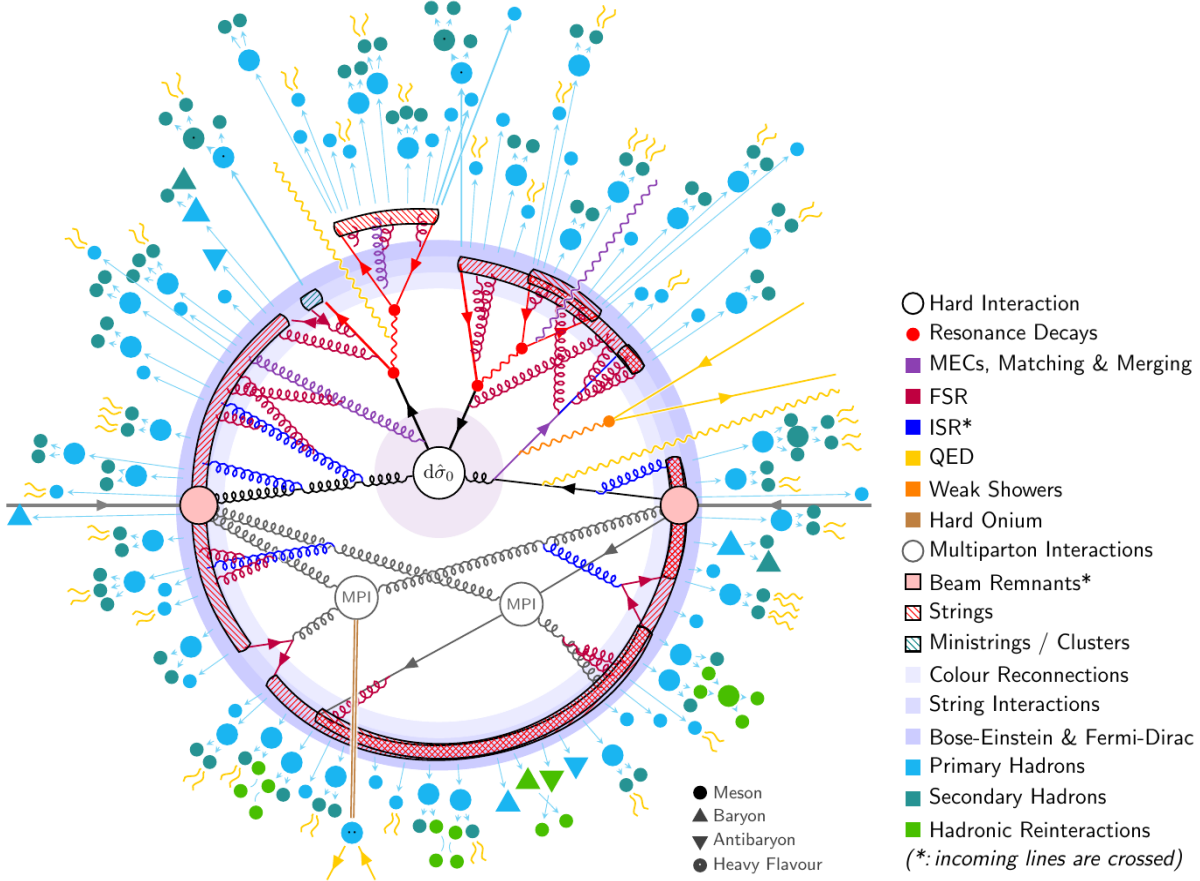
Figure 1.4: [29] Representation of a proton-proton collision where a top-antitop quark pair is produced and its subsequent PS. For the sake of clarity, simplifications have been made as detailed in the caption of Fig. 1 of [29].

We can write the cross section of an interaction in a proton-proton collision where one parton from one proton interacts with another parton from the other proton as follows:

$$\sigma_{2\to n} = \sum_{a,b} \int dx_a dx_b f_a(x_a, \mu_F) f_b(x_b, \mu_F) \hat{\sigma}_{ab\to n}(\hat{s}, \mu_F, \mu_R). \qquad (1.24)$$

In this expression, we sum over all possible partons that can give us the interaction we want at the order (in perturbation theory) we desire, and integrate over all the possible initial-state momenta of the partons, that are parameterised depending on the Bjorken variables $x_{\text{parton}} := \frac{p_{\text{parton}}}{p_{\text{proton}}}$. The functions $f(x_{\text{parton}}, \mu_F)$ are the PDF that depend, apart than on the momentum

of the interacting parton, on the factorisation scale μ_F. Finally, we have the cross section for the $ab \to n$ interaction at the desired order in perturbation theory, that depends on both the factorisation and renormalisation scales as well as on the energy of the interaction. This last term can be divided into

$$\hat{\sigma}_{2 \to n}(\hat{s}, \mu_F, \mu_R) = \frac{1}{2\hat{s}} \int d\Phi_n |\mathcal{M}_{ab \to n}(\hat{s}, \Phi_n, \mu_F, \mu_R)|^2, \tag{1.25}$$

where $\hat{s}$ is the s Mandelstam variable of the two partons, Φ_n refers to the entire phase space of the final-state particles n and $\mathcal{M}$ is the **matrix element** (ME) of the S-matrix (usually containing only the interacting parts of the matrix) that links the ab state with the n one up to the desired order in the perturbation series.

When modelling proton-proton interactions, the aftermath of the hard process should be considered. Quarks and gluons (partons) due to the colour confinement nature of the QCD interaction undergo a process in which they end up forming hadrons. This starts with a fragmentation where they radiate other partons, a perturbative QCD process, but due to the loss of energy of the final-state particles, it evolves into a non-perturbative process, where gluons and quarks end up conforming hadrons. The whole cascade of QCD radiations and emissions is called **parton shower** (PS).

Usually, some of the radiations from partons either in the initial or final states are considered inside of the ME. However, this adds difficulties to the ME calculations. The modelling of the whole process is thus separated between ME and PS, and the matching between both is taken into account experimentally. The radiation that is not considered in the ME that happens after the final-states particles are produced (final-state radiation, FSR), or before the initial ones perform the hard interaction (initial-state radiation, ISR) is modelled perturbatively up to a scale of energy $O(10^0 \text{GeV})$, starting from which a non-perturbatively approach is used for the rest of the process of hadronisation, where also other aspects, such as the decays the recently created hadrons might have (e.g. yielding leptons) before being measured in the detectors, are taken into account. Inside the parton shower, a large number of interactions between partons (or multiple partonic interactions, MPI) is used to model low-energy QCD processes. It has also been shown that the modelling of the colour charge flow between both protons is relevant to experimental observables and thus having an effect in the differential cross section or precision measurements of particle properties.

It should be noted that there are other partons inside the protons that might (or not) interact as well as the ab partons from the previous Eq. (1.24) and (1.25). The interactions that the rest of the partons can have among themselves are called the **underlying event** (UE) of the collision, and are taken into account experimentally.

Another aspect to take into account is the mass of initial-state b quarks. There are two approaches used by the community: to consider it as a massive quark, or to assume that it is massless. The latter option is called the five flavour scheme (5FS), because the bottom quark

is assumed as massless and a PDF is extracted for it inside the proton, as a parton more. The alternative is called four flavour scheme (4FS), where the b quark is massive and there are only four quark flavours (up, down, charm and strange) allowed in the proton. Both approaches, 4FS and 5FS, have been proven useful to model data and to do precision measurements, though differences can be found when, for example, measuring differential cross sections [30].

1.4 Top quark physics

The work done in this thesis is focused around the top quark: member of the third generation of leptons, and the most massive particle of the SM. The latest combined measurement of its mass, from LHC and Tevatron using the top-antitop quark pair production process, yields 172.69 ± 0.30 GeV [31], ≈ 50 GeV heavier than the Higgs boson, more than forty times heavier than the second heaviest quark, the bottom, and five orders of magnitude when comparing with the first generation quark masses (up, down). This mass value implies that its Yukawa coupling is almost the unity, being an exception for the rest of fermions. Because of this, it is usually said that the Yukawa coupling of the top quark is "natural", as the others are orders of magnitude smaller than one. The previous combination of quark masses has a total 0.17% total uncertainty. Individual measurements from experiments have been done at several centre-of-mass energies from both the Tevatron and the LHC in several decay channels, that have overall ≈ 1 GeV or less of uncertainty in all cases.

Another consequence of the top quark mass value is that it has a close relationship with the SSB and the Higgs boson, giving sensible corrections to its propagator. This attracts the attention from BSM proposals, that study the top quark properties, or similar particles, in order to provide answers for the hierarchy issues of the SM. Examples of such are searches for SUSY top squarks (or stops: the hypothetical supersymmetric partner of the top quark), that are expected to have a significant branching ratio to the top quark. Investigation of such particles has been done in ATLAS (e.g. [32, 33]) and also CMS (e.g. [34, 35]). Another BSM proposals might have new particles with a large mass, unreachable with current colliders. A proxy to them could be through the couplings between the top quark and other SM particles. These could be modified by BSM proposals, and deviations in them can be studied through effective field theories. Several articles from both ATLAS (e.g. [36, 37]) and CMS (e.g. [38, 39]) have such kind of interpretations.

The top quark mass affects many of its properties, in particular its decay width, that is consequently large ($1.42^{+0.19}_{-0.15}$ GeV [31]), implying a very small lifetime, of order $\tau \sim 10^{-25}$ s. This value is one order of magnitude smaller than the hadronisation time scale, $\sim 1/\Lambda_{\mathrm{QCD}} \sim 10^{-24}$, and this makes the top quark to decay in practically all cases even before having the chance to hadronise, which is a special feature that does not happen with the rest of the quarks. Thanks to it, we can study its decay products and derive properties from the free top quark itself, such as asymmetries, polarisation, or decay width. This cannot be done experimentally with other

final-state quarks, as they hadronise to conform hadrons (and thus the free particle information is convoluted with that of other quarks, effectively loosing it).

One example of top quark properties is polarisation of the top quark, or of its decay particles such as the W boson. Probes of such can be obtained from angular distributions, as done by ATLAS (e.g. [40]) or CMS (e.g. [30]). Another one is the forward-backward asymmetries in top-antitop quark pair production, checked in the Tevatron [41], as in higher orders in QCD is expected to arise. In the LHC its measurement is harder due to the proton-proton collisions, instead of proton-antiproton. Thus, a charge asymmetry is defined depending on rapidity, with measurements from both ATLAS and CMS [42].

1.4.1 Decay channels

Although various channels could a priori be studied for the top quark decays, in practice almost only one should be considered: the decay to a W boson and a bottom quark. This is explained when checking the observed values of the CKM matrix that affect the top quark [31],

$$|V_{td}| = (8.0 \pm 0.3) \cdot 10^{-3}, \quad |V_{ts}| = (38.8 \pm 1.1) \cdot 10^{-3}, \quad |V_{tb}| = (1.013 \pm 0.030), \quad (1.26)$$

clearly indicating that the most favoured decay is that where the final-state quark is its companion in the third generation of leptons. Consequently, the decay of the W boson determines the decay categorisation of the top quark production processes. These can produce leptons in the final state (leptonic channel) or more quarks (hadronic channel). The sum of all leptonic ($e\nu_e$, $\mu\nu_\mu$, $\tau\nu_\tau$) decay channels is roughly a third of the total ones, being the remaining all hadronic. Among the leptonic ones, their decay probabilities are roughly the same for the three possibilities.

1.4.2 Production modes

Top quark processes in hadron-hadron collisions are usually grouped into three sets, depending on the other final-state particles, and the production cross section.

1.4.2.1 Top-antitop pair

By far, the most important way of producing top quarks in hadron colliders is through the production of a top-antitop pair, or $t\bar{t}$. The leading-order (LO) Feynman diagrams that describe this production mode can be seen in Fig. 1.5. In proton-proton colliders, the most relevant diagram is the gluon fusion gg channel, whereas in proton-antiproton colliders, the quark-antiquark $q\bar{q}$ dominates. The most precise prediction of its cross section for a centre-of-mass energy of $\sqrt{s} = 13$ TeV yields $\sigma_{t\bar{t}} = 832^{+46}_{-51}$ pb, when considering a top quark mass of 172.5 GeV. The calculation is at NNLO in QCD, resumming the next-to-next-to-leading logarithmic (NNLL) soft gluon terms with top++2.0 [43, 44, 45, 46, 47, 48, 49]. The PDF and α_S uncertainties are considered in the

prediction, obtained following the PDF4LHC recommendations [50] with the MSTW2008 68% CL NNLO [51, 52], CT10 NNLO [53, 54] and NNPDF2.3 5f FFN [55] PDF sets. This source is added in quadrature to the scale uncertainty as the effect of varying the top quark mass by ± 1 GeV.

This process has been measured various times in different decay channels and energies. The D0 and CDF Collaborations at the Tevatron (first collider where $t\bar{t}$ was observed), considered both the dileptonic [56] as well as semileptonic [57, 58] channels to measure it. Recently, a combination from both ATLAS and CMS released the combination of their measurements of this process at $\sqrt{s} = 7$ and 8 TeV [59]. In a lower energy, $\sqrt{s} = 5.02$ TeV it has been also observed inclusively by ATLAS [60] and CMS [61]. Dedicated inclusive measurements are available also at 13 TeV (e.g. [62, 63]) and more recently, at $\sqrt{s} = 13.6$ TeV [64, 65]. The process has been studied differentially in several occasions too (e.g. [66, 67]).
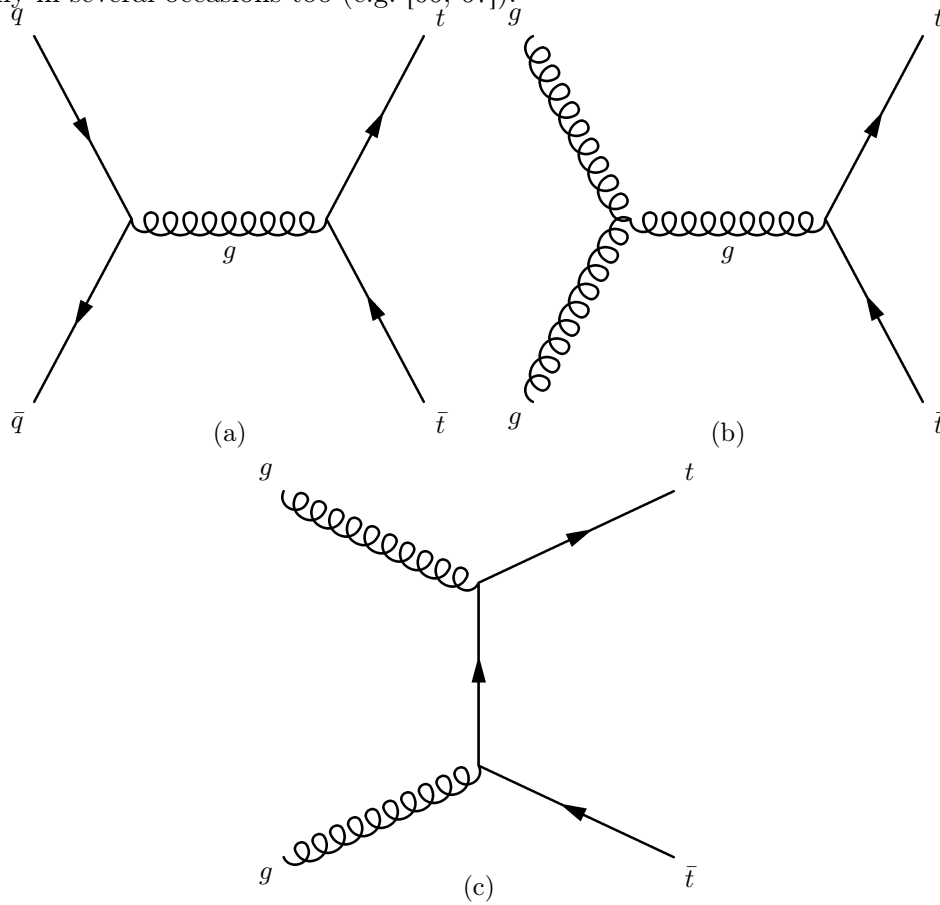
Figure 1.5: Leading-order Feynman diagrams for $t\bar{t}$ production. (a) shows the quark fusion process, whereas (b) and (c) represent the gluon fusion production modes.

As the top decays in almost all cases to a W boson and a bottom quark, the decay channels of the $t\bar{t}$ process are classified depending on the subsequent decay of the boson, as experimentally speaking the quark will hadronise after being produced. Thus, we speak of dileptonic decay channels, where both $W^{\pm}$ bosons decay to only leptons, i.e. they go to $\ell_1\nu_{\ell_1}\ell_2\nu_{\ell_2}$. If one of those

W bosons decays to quarks we are describing semileptonic production modes, $\ell\nu_\ell q\overline{q}'$. Finally, if both $W^\pm$ decay to quarks, we call those final states hadronic, $q_1\overline{q}'_1 q_2\overline{q}'_2$. The expected decay proportions, according to the Particle Data Group [31] are shown in 1.6.
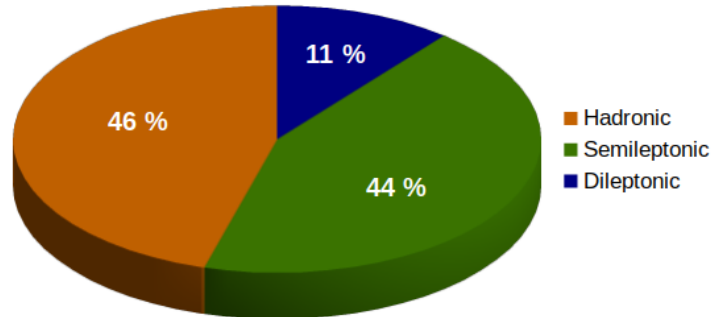
Figure 1.6: Graph showing the proportions of the decay modes of the $t\overline{t}$ process.

1.4.2.2 Single-top

The second group in terms of production cross section (at proton-proton colliders) is the set of processes where only one top quark is produced in their final state and is called "single top". There are three processes, characterised depending on the role of the W boson in the Feynman diagram, using the Mandelstam variables: in the t and s channels the boson plays an intermediate part, whereas in the tW channel it belongs to the final state.

Whereas pair production features in their LO diagrams only QCD interaction vertexes, single top processes contain also an electroweak vertex, allowing to experimentally probe the CKM entries. In addition, the resulting top quark is polarised due to the presence of this vertex. The cross section of these processes is lower than pair production, but is still significant. The LO Feynman diagrams of the three processes are shown in Fig. 1.7.

t-channel This is the first process, in terms of production cross section, of the single top group. The most precise prediction of its cross section (including both the top production process and the antitop one) for a centre-of-mass energy of $\sqrt{s} = 13$ TeV in proton-proton collisions yields $\sigma_{t-ch.} = 214.2^{+2.4}_{-1.7}\,(\text{scale})^{+3.3}_{-2.0}\,(\text{PDF},\alpha_S)\,\text{pb} = 214.2^{+4.1}_{-2.6}\,\text{pb}$ [68]. This value is calculated with the `MCFM` program at NNLO in QCD using a top-quark mass of 172.5 GeV. The quoted uncertainties include the uncertainties due to the choice of μ_r and μ_f, the uncertainty in the PDFs and in the value of α_s. The scale uncertainty is determined by varying μ_r and μ_f independently up and down by a factor of two, whilst never allowing them to differ by a factor greater then two from each other. The combined PDF and α_s uncertainties were determined according to the PDF4LHC21 PDF set prescription [69].
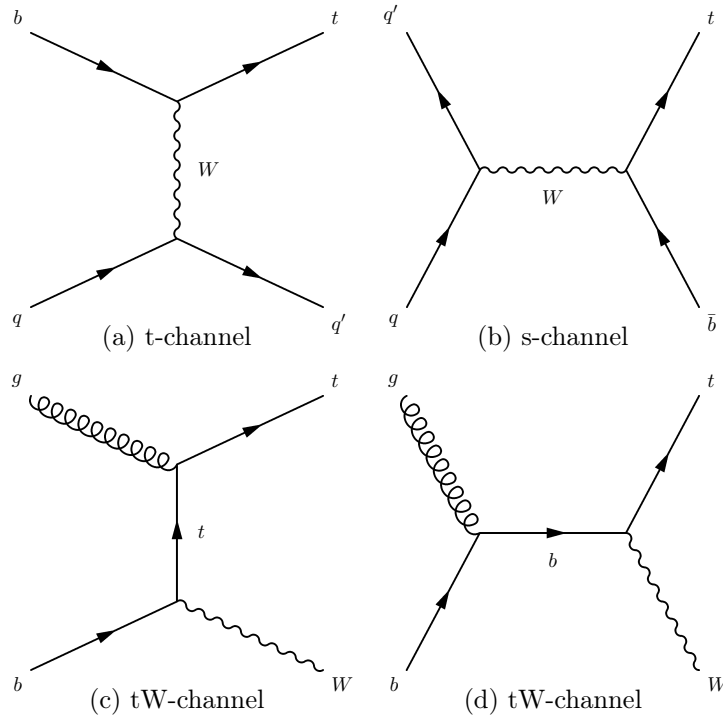
Figure 1.7: Leading-order Feynman diagrams for single top processes.

Experimentally, this process is characterised by the decay modes of the only top quark, classified in a similar way as with the top-antitop pair production: leptonic ($\approx 33\%$), or hadronic ($\approx 66\%$). In addition, its signature has a softer (in terms of energy) jet, coming from the quark (usually called "spectator" quark) partner with the top in the final state. This jet has usually a momentum with low angles with respect to the proton beam axis.

Studies done in the t-channel can explore the ratio between the production of top and antitop quarks in proton-proton collisions, that is directly related with the PDFs of those hadrons. This is also the reason behind the difference in its cross section values when comparing the top and the antitop production. Analysis can also benefit from reconstructing experimentally the top quark, and describe its properties, such as the spin asymmetry.

This process was first observed by the D0 and CDF collaborations at the Tevatron, in proton-antiproton collisions [70, 71]. In the LHC, both ATLAS and CMS have measured its inclusive cross section at $\sqrt{s} = 7$, 8 and 13 TeV [72, 73, 74, 75, 76, 77], as well as its differential cross section at the same energies [76, 73, 78, 30]. Some of these analyses exploit the data to study properties of the top, such as its polarisation [30, 79] or probes to the value of $|V_{tb}|$ [73, 80].

tW-channel The tW channel has the second largest cross section in proton-proton collisions. The prediction at a centre-of-mass energy of $\sqrt{s} = 13$ TeV is of $\sigma_{\text{tW-ch.}} = 71.7 \pm 1.8 \,(\text{scale}) \pm 3.4 \,(\text{PDF},\alpha_S)\,\text{pb} = 71.7 \pm 3.8\,\text{pb}$ [81, 82], with an aNNLO+NNLL prediction in QCD. This estimation has been obtained with the MSTW2008 68% CL NLO [52, 51] PDF dataset and

assuming a mass of 172.5 GeV for the top quark. A recent update of this prediction has been given recently, with aN^{3}LO precision in QCD and resummations at NNLL. This new prediction is $\sigma_{\text{tW-ch.}} = 79.3^{+1.9}_{-1.8}$ (scale) ± 2.2 (PDF,α_S) pb $= 79.3^{+2.9}_{-2.8}$ pb [83]. The same 172.5 GeV were assumed as mass for the top quark, but using in this case the PDF4LHC21 PDF set [69].

Any analysis done in the tW channel must face the same challenge: the close similarity of this process with $t\bar{t}$ production (a connection explained in detail in section 1.4.3). Experimentally, it is a challenge to separate collisions where this process has occurred from those where the pair production happened. Thus, measurements of tW depend crucially on this separation and the estimation of the effects that $t\bar{t}$ has in the extraction of tW.

The tW channel was not accessible at the Tevatron due to its small cross section in proton-antiproton collisions at $\sqrt{s} = 1.96$ TeV. At the LHC, however, evidence for this process in 7 TeV collision data was presented by ATLAS [84] and CMS [85]. At $\sqrt{s} = 8$ TeV, measurements by CMS [86] and ATLAS [87] were in good agreement with theoretical predictions as well as at 7 TeV.

s-channel The single top process with less production cross section is the s-channel. It has, at $\sqrt{s} = 13$ TeV and proton-proton collisions, a total (top and anti-top productions) cross section of $10.32^{+0.29}_{-0.24}$ (scale) ± 0.27 (PDF,α_S) pb $= 10.32^{+0.40}_{-0.36}$ pb, seven times smaller than tW's. The s-channel is particularly sensitive to the PDFs of the colliding hadrons, and because of that it has already been observed at the Tevatron (proton-antiproton collisions) [88] and not yet at the LHC (proton-proton). There have been nevertheless searches of this process at $\sqrt{s} = 7$, 8 and 13 TeV from both ATLAS and CMS [89, 90, 91, 92], and some of them obtaining evidence of it [91, 92].

1.4.2.3 Others

The remaining top quark production processes, with smaller cross section in proton-proton collisions, are not always categorised in the same group. However, in this document we divide them by a notation commonly used. Those processes where top quarks in the final state are accompanied with other vector bosons are called **associated production** modes.

Most of them have a top-antitop quark pair in their final state, with a boson: $t\bar{t}Z$, $t\bar{t}W$, $t\bar{t}\gamma$ and $t\bar{t}H$. There are also "single-top associated production" processes: although tW in principle could be considered one of them, it is usually classified inside the single top category. To the processes mentioned before, we can add others such as tZq, $t\gamma$, tWZ, tH or tHW.

There are other processes that cannot be considered as associated production (or other category), or that have a very small production cross section. Inside this category, we could mention the $t\bar{t}t\bar{t}$ process, that has been recently studied by both ATLAS and CMS [93, 94].

1.4.3 The tW and t$\bar{t}$ relationship

When experimentally studying the tW process at NLO in QCD, one can find that it has final states that are shared with t$\bar{t}$ (namely, the *tWb* final state). Given the theoretical framework to simulate these processes (the S-matrix formalism), strictly speaking we cannot use, at NLO in QCD, tW and t$\bar{t}$ as completed separate processes because processes are characterised by their final states. Therefore, one should consider a *tWb* final state, or more commonly (to include off-shell effects in the tops also), the WWbb final state, and calculate its matrix element to simulate it. This has been accomplished for some decay channels already of the WWbb process [95].

These new samples are still being initially studied, and they are not widely used in the community: most of the analyses use separate tW and t$\bar{t}$ Monte Carlo simulations (MC) at NLO in QCD. This poses a couple of problems, however. First of all, there will be events that correspond to Feynman diagrams that can be in both samples, and thus they will be double counted if both samples are used simultaneously.

Secondly, we are ignoring the effect of the interferences between the Feynman diagrams that appear when calculating the squared amplitude of the matrix element. In tW and t$\bar{t}$ at NLO in QCD, Feynman diagrams that can have a top-antitop pair on-shell are called "doubly resonant" diagrams (and come from the t$\bar{t}$ process), whereas those where only one can be on-shell are called "singly resonant" (and come from tW). Examples of the first set can be seen in Fig. 1.8. We can thus write the entry of the S-matrix for WWbb as
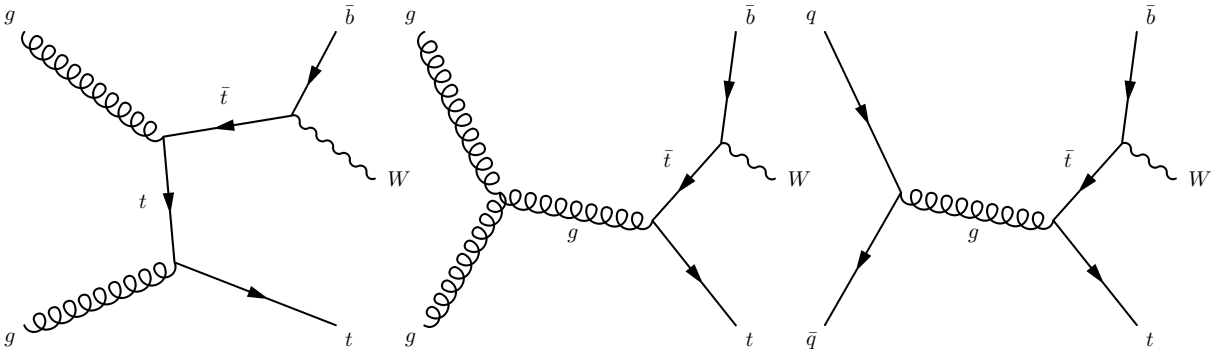
Figure 1.8: Examples of Feynman diagrams for the *tWb* final state achievable when adding in perturbation theory one extra QCD order to the tW process.

$$\mathcal{M}_{\text{WWbb}} = \mathcal{M}_{\text{doubly res.}} + \mathcal{M}_{\text{singly res.}} + \mathcal{M}_{\text{others}}, \quad (1.27)$$

where we have put inside $\mathcal{M}_{\text{others}}$ the rest of Feynman diagrams that can yield the WWbb final state [95]. When calculating its squared amplitude, we get, in particular:

$$|\mathcal{M}_{\text{doubly res.}}|^2 + |\mathcal{M}_{\text{singly res.}}|^2 + 2 \cdot \text{Re}\left(\mathcal{M}^*_{\text{doubly res.}} \cdot \mathcal{M}_{\text{singly res.}}\right). \quad (1.28)$$

The last term is the quantum interference between the singly and doubly resonant Feynman diagrams. If one uses separate tW and t$\bar{t}$ samples, these are neglected (as well as other residual

contributions from $+\mathcal{M}_{\text{others}}$, and their interferences also).

To avoid or reduce the effect of these two issues when using the separate samples, one can alter them: there are two general approaches in the field that modify the tW process MC sample to accomplish this. The first one is the diagram removal (or DR), where all the doubly resonant diagrams are removed from the matrix element calculation. This avoids entirely the double counting, though it doesn't take into account the interference terms in the calculation. A deviation of this method, called DR2, adds these interferences to the calculus. Both DR and DR2 loose gauge invariance in the calculation, but this ends up having a small experimental effect in the modelling of the resulting tW MC samples.

The second approach, diagram subtraction (DS), does not modify the Feynman diagrams of the matrix element calculation, but it adds an artificial term to it, designed so that the cross section of the process tends to zero whenever we have two tops on-shell (or near on-shell). The calculus of squared amplitude has then in particular these terms,

$$\left|\mathcal{M}_{\text{doubly res.}} + \mathcal{M}_{\text{singly res.}}\right|^2 - \mathcal{S}, \tag{1.29}$$

where $\mathcal{S}$ is the new artificial term constructed so that it cancels exactly $\left|\mathcal{M}_{\text{doubly res.}}\right|$ when there are two top quarks on-shell, is gauge invariant, and tends to zero when we leave the surrounding area to the top quark pole mass in the phase space. A generic subtraction term $\mathcal{S}$ can be written thus as

$$\mathcal{S} = \mathcal{S}\left(\{p_i\}\right) = f\left((p_W + p_b)^2\right) \left|\mathcal{M}_{\text{doubly res.}}\left(\{q_i\}\right)\right|^2, \tag{1.30}$$

where $\{p_i\}$ is the four-momenta of the initial and final-state particles and $\{q_i\}$ are the same four-momenta after a basis change that makes one of the top quark four momenta (the internal one, from doubly-resonant diagrams with tWb final state) always on-shell (i.e. $\{q_i\} : (q_W + q_b)^2 = m_t^2$). Thanks to this change in basis (that is not a Lorentz transformation) for the q_i four-momenta, we can guarantee gauge invariance of $\mathcal{S}$ and thus of the entire matrix element in the $\Gamma_t \to 0$ limit. Later, in practice, non-zero Γ_t are achievable with no appreciable experimental effect in MC simulations.

There are variations on how to define $f\left((p_W + p_b)^2\right)$, that partly characterise the different DS models. The most direct one uses a quotient of Breit-Wigner distributions (the top quark before and after the basis change),

$$f_1(s) := \frac{(m_t\Gamma_t)^2}{(s - m_t^2)^2 + (m_t\Gamma_t)^2}, \tag{1.31}$$

and is the one most commonly used. It is implemented in both `POWHEG` and `MadGraph5_aMC@NLO` generators. This variation is usually called DS alone, or DS1. Another way of defining f is modifying f_1, and changing $m_t\Gamma_t$ by $\sqrt{s}\Gamma_t$,

$$f_2(s) := \frac{(\sqrt{s}\Gamma_t)^2}{(s - m_t^2)^2 + (\sqrt{s}\Gamma_t)^2}. \tag{1.32}$$

This variation introduces a fundamental difference with f_1, as the resonance of the Breit-Wigner now can depend on the process. This approach is usually called DS2.

Apart from defining f, there is another degree of freedom in the diagram subtraction method: the reshuffling of the four-momenta. In the two most common generators, `POWHEG` and `MadGraph5_aMC@NLO`, it is configured to be done with all the initial & final-state particles. `MadGraph5_aMC@NLO` adds another possibility: to do it by only changing the initial-state particles four-momenta, adding an indirect dependence on the PDF used for them. Thus, in practice, we have four DS approaches: DS1 with the reshuffling done with all initial & final-state particles, and only with the initial ones, and analogously for DS2.

When using these combined separated samples, a common practice in the field is to take into account this artificial modification of the tW sample by adding an extra source of uncertainty in the analyses, consisting usually in the difference between the MC expectations for the DR and DS(1) approaches. Although usually analyses dedicated to the tW process or in a closer phase-space take into account the interference with $t\bar{t}$ and design the study to minimise its effects, in some investigations one can see that this uncertainty source is one of the most important ones for some signal regions [96].

Getting rid of this extra uncertainty is one of the reasons to work directly with the WWbb final state, that has an already mentioned sample of some of its final states: those dileptonic with opposite flavour leptons (hence its "other" naming: `bb4l`). This sample is done in the 4FS, includes all interference terms between tW and $t\bar{t}$ by construction, and also has some improvements over other commonly used samples for the separate processes: a consistent NLO+PS treatment of top resonances (including quantum corrections to top propagators, and off-shell top-decay chains), and also a complete NLO accuracy in both ME production and also decays (including spin correlations, and interference between NLO radiation from top production). The ATLAS Collaboration has published an early study with unfolded data [97] and also extensive generator level studies [98].

2 Experimental setup

THE 16$^{\text{th}}$ of December of 1994, the CERN Council approved the construction of a proton-proton (also lead-proton and lead-lead) particle collider in a previous electron-positron accelerator, the old LEP's (Large Electron Positron) tunnel [99]. The collider, named Large Hadron Collider (LHC) was envisioned to achieve energies in the centre-of-mass frame of $\sqrt{s} = 14$ TeV, never reached before. The potential measurements done under such conditions could allow significant scientific discoveries, such as the observation of the Higgs boson. The year after, a budget of order ≈ 3000 M€ was approved, although the construction works finished a decade later, in 2008.

The data used in this thesis have been recorded by the Compact Muon Solenoid (CMS) experiment, that studies collisions provided by the LHC. Therefore, we will describe in this chapter both, beginning with an overview of the particle collider, and then going in a detailed view of the CMS detector. Afterwards, the reconstruction of particle objects from collisions in CMS is covered, followed by a summarised view of the future of the LHC and the studies that were done to prepare the CMS detector for it. The final section is an overview of a generic HEP analysis, provided to guide the next chapters of the document.

2.1 The Large Hadron Collider

The LHC has circular shape, installed in a tunnel with roughly 27 km of length that is located near the Swiss city of Geneva and crosses the border (two times) with France. This is the area where CERN's headquarters are located (in the village/suburb of Meyrin). The collider is slightly tilted with respect to the surface, roughly being around 100 m underground. This is not casual, as it provides protection against cosmic rays as well as isolation from nearby vibrations.

As it was mentioned, it was put in the same tunnel that the old Large Electron Positron (LEP) collider was. This was done for economic reasons, as well as for convenience: CERN has a large amount of accelerators, colliders, and other scientific experiments in the area. The LHC benefits from this, mainly using some of these installations to pre-energise the colliding particles. Then, the cations travel inside two separate tubes (or "pipes", in LHC's argot) in opposite directions. These tubes cross each other in various locations, and there is where collisions happen, and where the main detectors have been constructed to study them: ATLAS (A Thoroidal LHC Apparatus), ALICE (A Large Ion Collider Experiment), CMS and LHCb (LHC-beauty). While ALICE and

LHCb are designed to be specialised in some investigations, ATLAS and CMS have a general-purpose outline.

The design energy (in the centre-of-mass reference frame) of the collider is $\sqrt{s} = 14$ TeV in proton-proton collisions, obtained by accelerating them to 7 TeV each. However, that value has not been reached yet. It has been working at lower settings, such as $\sqrt{s} = 7, 8$ or 13 TeV, and currently works at $\sqrt{s} = 13.6$ TeV.

2.1.1 Running layout

The working procedure of the LHC requires of other accelerators that energise previously the cations that are going to collide. We will show here the proton-proton collision workflow, as it is the one of interest for this thesis. The ion-proton, or ion-ion is slightly different at its beginning, as well as in the maximum energies the particles can achieve.
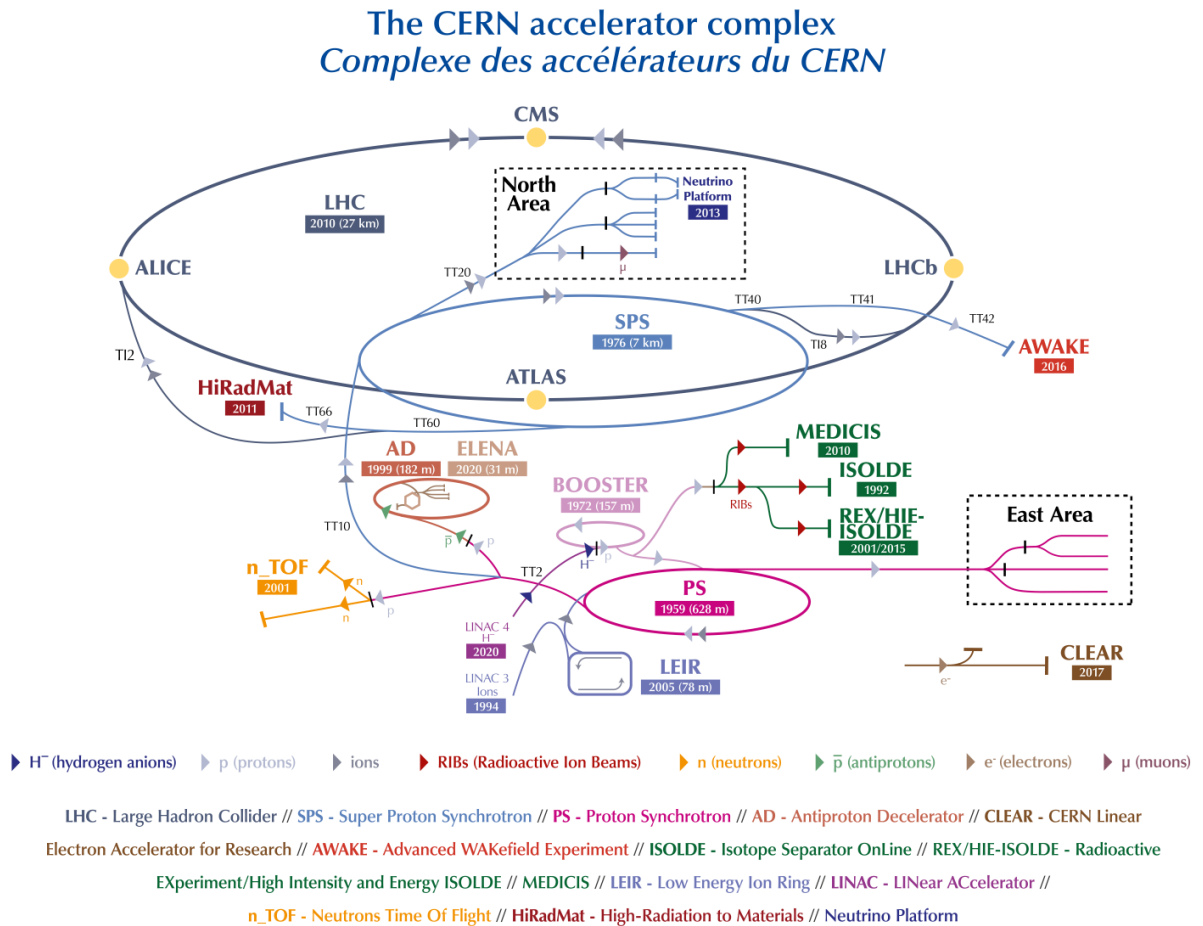
Figure 2.1: [100] Diagram showing CERN's accelerator complex.

An outline of the whole CERN's set of accelerators and their connections can be seen in Fig. 2.1.

To run the LHC, we do not use all of them, but some: the journey begins with a H_2 bottle, as protons are obtained by ionising hydrogen atoms. Then, negative ionised hydrogen atoms, H^- (i.e. one proton with two electrons) are passed to a linear accelerator, LINAC4, that energises them up to 160 MeV. They are then separated into protons and electrons. Protons are handled to a circular synchrotron, the Proton Synchrotron Booster (PSB), that accelerates them up to 2 GeV. This is the current workflow, but the data used in this thesis was obtained with the older LINAC2, that only energised protons directly (not H^-) up to 50 MeV, before going into the PSB. Later, protons are injected into the Proton Synchrotron (PS) where they acquire an energy of 25 GeV. Afterwards, they are passed onto the Super Proton Synchrotron (SPS), where they go up to 450 GeV before entering, in opposite directions and pipes, the LHC. There, the final acceleration is done up to the collision energy. Protons travel the LHC in bunches of $\approx 10^{11} p^+$, separated temporally by 25 ns.

The whole process of activating the complete set of accelerators and gradually increase proton's energy is extremely complex. It requires several people ("operators") that must work continuously, day and night, in shifts, to ensure that the complete procedure works as expected, including being in synchronisation with those responsible of the experiments where the particles are expected to arrive (and collide or interact). This work, that not always has the visibility that corresponds to its importance, is essential to get scientific results (such as those of this thesis) that usually arrive to journals much later in the form of an article.

2.1.2 Essential components

To accelerate particles the LHC has eight radiofrequency cavities formed of superconductor magnets that work at 4.5 K of temperature, using liquid helium as refrigerator. By the means of a precise control of the oscillation frequency, particles travelling suffer an induced electrical field that accelerates them. This, apart from energising them, allows to ensure a correct separation of the bunches travelling, as they must be in synchronisation with the oscillation frequency. This is crucial to ensure also that the bunches cross each other in the designed places to do so. A diagram of an LHC dipole can be seen in Fig. 2.2.

Thousands of magnets are used to make the LHC work. From them, the most outstanding are the 1232 supercontuctor dipoles, that are dedicated to force hadrons to turn. They offer 8.33 T of magnetic field while working at 1.9 K of temperature, cooled thanks to liquid helium. In addition, multipole magnets are used to control the geometrical spread of the beams. This allows a larger rate of collisions in most of the experiments obtained by focusing the bunches (or a lower one, such as in LHCb, by intentionally collimating them less).

2.1.3 Luminosity and pileup

Although we have dedicated an entire chapter to explain the theoretical prolegomena needed to understand this work, and we just briefly explained how the LHC provides us with collisions
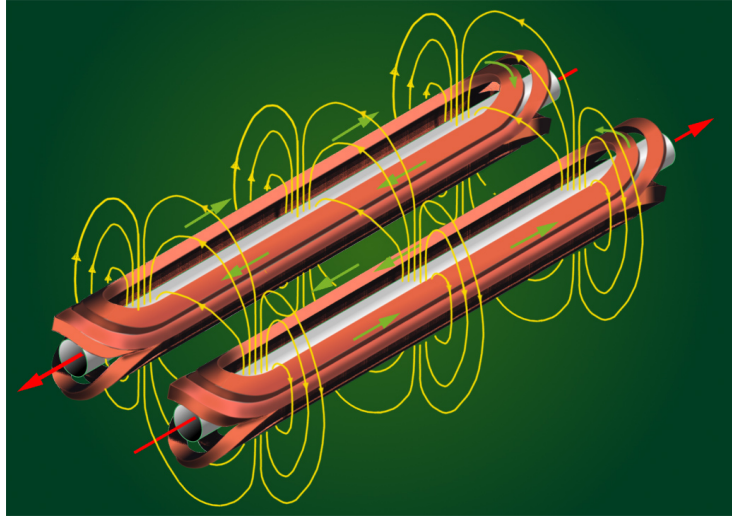
Figure 2.2: [101] Diagram showing an LHC radiofrequency cavity. Trajectory of travelling particles is shown with red arrows, while green ones indicate the direction of the electric field. Yellow lines represent the generated magnetic field.

among hadrons, we still need a "bridge" among the expectations and the data that we can collect with our detectors. This link is the relationship between the collisions rate and the cross section of one process, that we can express differentially as

$$\frac{dN}{dt} = \mathcal{L} \cdot \sigma, \tag{2.1}$$

where $\mathcal{L}$ is called **instantaneously luminosity**. It can be directly seen that it has the inverse dimensions of the cross section as well as one over time, i.e. one over area times time. This equation relates the theory we have (instantiated as its cross section) with the experimental outcomes (the direct counts of the interactions), allowing us to confront the collected data with the predictions from e.g. the SM (or any other theory) by adding a subscript to both N and σ (e.g. proton-proton collisions, or tW). The luminosity depends only on the experimental collider setup, i.e. the LHC settings, and gives us a notion of the performance of the machine in terms of collisions per time with respect to any reference collision process. The equity can be also written in absolute terms,

$$N = L \cdot \sigma, \tag{2.2}$$

having now the **integrated luminosity** L. This trivially is defined as $L := \int \mathcal{L} dt$, has only units of inverse area, and provides us a reference of the amount of collisions that have happened over a given period of time. The LHC design luminosity was $\mathcal{L} = 1 \cdot 10^{-34} \mathrm{cm}^{-2} \mathrm{s}^{-1}$, though now it this nominal value has been surpassed and $\approx 2 \cdot 10^{-34} \mathrm{cm}^{-2} \mathrm{s}^{-1}$ has been achieved. In Fig. 2.3, the luminosity (in cumulative terms) that the LHC provides to the CMS experiment during its stable collision conditions (commonly called "stable beams") for proton-proton collisions is shown

for all data-taking periods of the installation, or "Runs": Run 1, Run 2, and Run 3 (up to the year 2022).
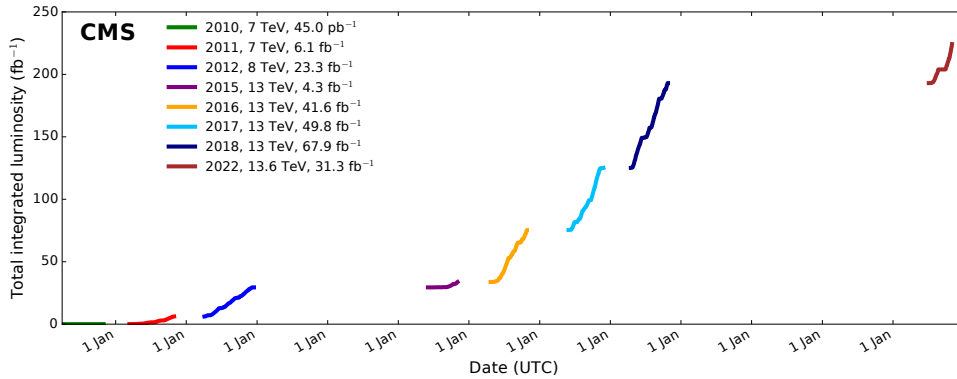
Figure 2.3: Cumulative luminosity versus time delivered to CMS in stable conditions from all data-taking periods of the LHC: Run 1, Run 2, and Run 3 (up to 2022) of proton-proton collisions.

The luminosity that the LHC delivers to its experiments is not exactly the same as the one they record, as usually there are "dead times" where no data is collected. Therefore, deviations between these two values can be observed in plots such as those of Fig. 2.4, where the evolution of the total integrated luminosity over time is shown.
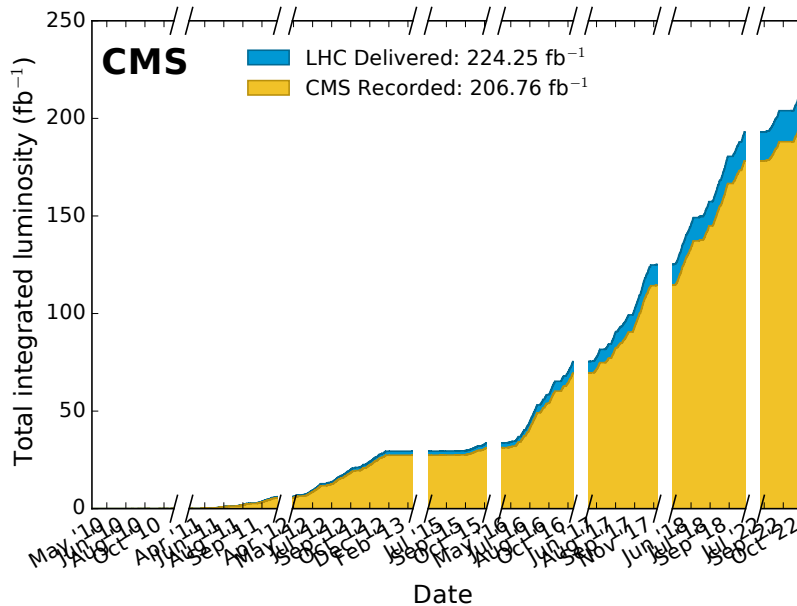
Figure 2.4: Cumulative luminosity versus time delivered and recorded by CMS in stable conditions from all data-taking periods of the LHC: Run 1, Run 2, and Run 3 (up to 2022) of proton-proton collisions.

There is another phenomenon that affects the collisions that the LHC provide to the experiments. One of the ways of achieving a large instantaneous luminosity is to "squeeze" the proton beams to reduce the collision area in the experiments. For colliding bunches of particles, the instantaneous luminosity $\mathcal{L}$ can be written as

$$\mathcal{L} = \frac{N_1 N_2 \nu}{A_{\text{eff}}}, \tag{2.3}$$

where A_{eff} is the effective area where the collisions take place, N_i the number of protons in each colliding bunch, and ν the collision frequency. This general expression must be corrected by various effects that modify the actual luminosity, such as the crossing angle (the bunches do not collide head-on). Taking into account these effects, the expression ends up being

$$\mathcal{L} = \frac{n_b N_1 N_2 \nu}{S} \cdot X, \tag{2.4}$$

where n_b is the number of bunches of protons, and S (with area units) and X (adimensional) depend on the shape of the beams and their relative inclination and offsets: "squezeed" beams reduce the area S, forcing a larger luminosity, and the opposite reduces it. A quick view on how these effects modify the instantaneous luminosity can be seen in [102], and a more detailed overview of luminosity calculations in proton-proton colliders in [103].

The collision of bunches instead of individual particles has implies that we obtain also a larger number of simultaneous collisions of protons between the bunches. When we record one collision datum, we focus our studies to only one of those interactions: that with the largest fraction of transversal energy, as our detector is designed to measure best in the transversal plane. The other collisions do not disappear, and they act effectively as background to our studies: we call this "pollution" (the number of additional interactions) **pileup**. In Fig. 2.5 a representation of a real collision that took place in 2016 and was collected by the CMS experiment is shown.

That event was collected under high pileup settings by the LHC, with roughly 100 simultaneous collisions in average [104]. The pileup depends on the collider setup, and it has changed over time with the evolution of the techniques used to run it. In Fig. 2.6 this variation can be seen for the data-taking periods Run 1 and Run 2. For the latter, the total average yields 34 as mean number of interactions per crossing.
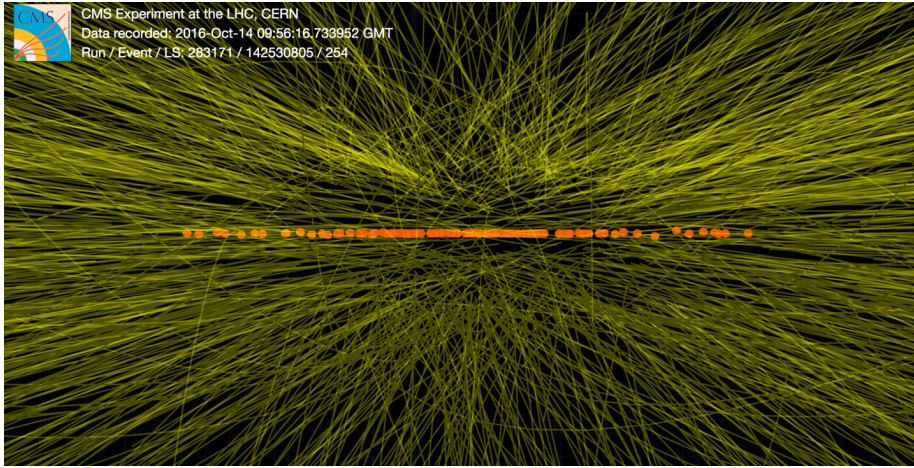
Figure 2.5: [104] Representation of a real collision taken under high pileup conditions by the CMS Collaboration in 2016. The yellow lines are the tracks of reconstructed particles emerging from the multiple interaction vertexes that can be identified (orange points).
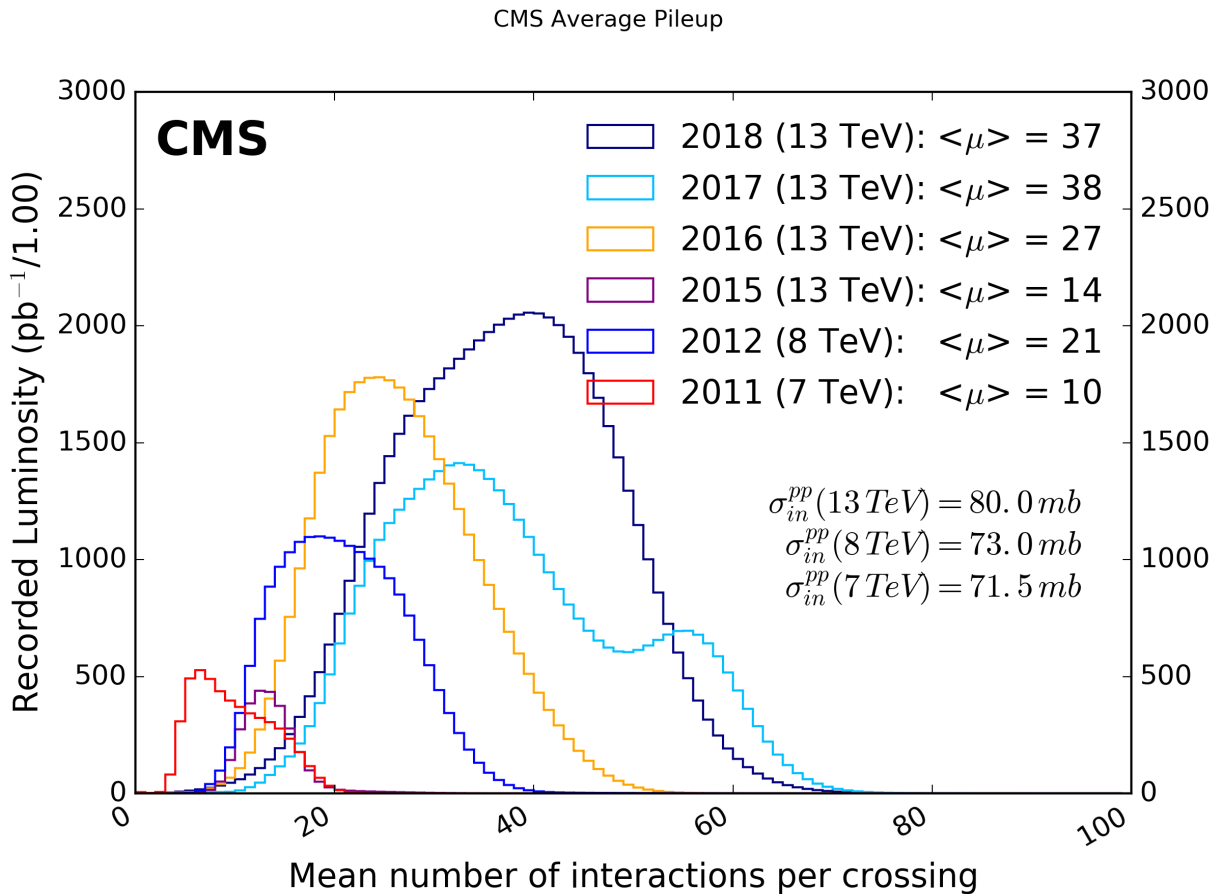
Figure 2.6: Evolution of the recorded luminosity by the CMS experiment against the mean number of interactions per crossing (i.e. average pileup) for the entire Run 1 and 2.

35

2.2 The CMS experiment

The Compact Muon Solenoid or CMS [105] is, as we said, one of the main detectors at the LHC and one of the two with a general purpose, along ATLAS. It is dimensionally speaking large and roughly with barrel or cylindrical shape, with 21.6 m of longitude and a diameter of 15 m, and it is made up of several subdetectors that are organised in cylindrical layers around the two tubes of the LHC that converge in its centre and cross it through its axis. It is structurally divided (Fig. 2.7) in barrel (and this one, subdivided in five wheels) and endcaps (which would "close" the barrel). Its main subdetectors, which will be briefly described later, are the detector of trajectories and collision vertexes or **tracker**, the **electromagnetic calorimeter** or ECAL, the **hadronic calorimeter** or HCAL, and the **muon system**. Between the HCAL and muon subdetectors, a large superconductor solenoid lies.

CMS DETECTOR

Total weight	: 14,000 tonnes
Overall diameter	: 15.0 m
Overall length	: 28.7 m
Magnetic field	: 3.8 T

STEEL RETURN YOKE
12,500 tonnes

SILICON TRACKERS
Pixel (100x150 μm²) ~1.9 m² ~124M channels
Microstrips (80–180 μm) ~200 m² ~9.6M channels

SUPERCONDUCTING SOLENOID
Niobium titanium coil carrying ~18,000 A

MUON CHAMBERS
Barrel: 250 Drift Tube, 480 Resistive Plate Chambers
Endcaps: 540 Cathode Strip, 576 Resistive Plate Chambers

PRESHOWER
Silicon strips ~16 m² ~137,000 channels

FORWARD CALORIMETER
Steel + Quartz fibres ~2,000 Channels

CRYSTAL ELECTROMAGNETIC CALORIMETER (ECAL)
~76,000 scintillating PbWO₄ crystals

HADRON CALORIMETER (HCAL)
Brass + Plastic scintillator ~7,000 channels

Figure 2.7: [106] Schema of the CMS detector as well as its subdetectors.

When particles cross the different components, they can be identified and characterised by the different responses they give in each subdetector, as Fig. 2.8 shows, and we will detail later in Sect. 2.3. These responses consist on electric signals collected by the different subdetectors. The mechanism through which these signals are generated depend on the technology each CMS components use to detect particles, and exploit the different ways these can interact with matter.
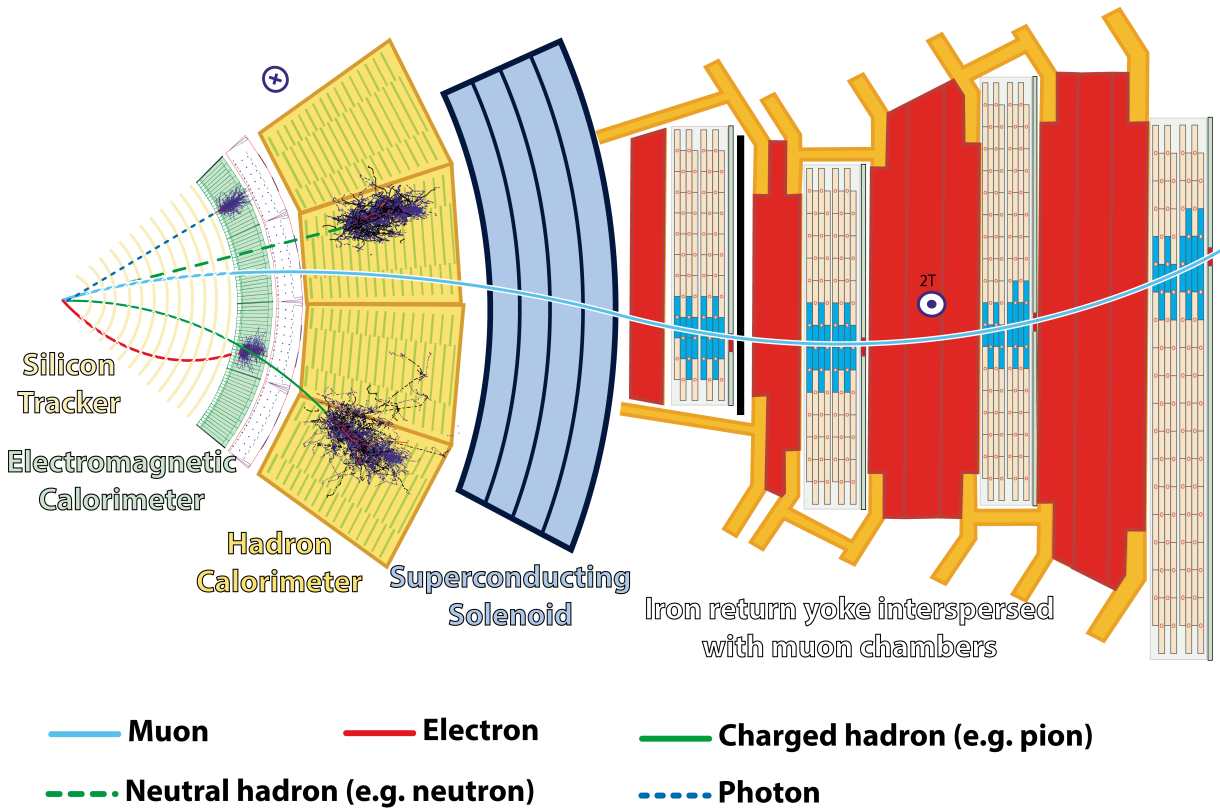
Figure 2.8: [107] Diagram of a circular sector of the detector, in which it can be seen how different particles go through it.

When considering e.g. how photons interact with matter, the Rayleigh scattering, photoelectric, Compton scattering and electron-positron pair production (in increasing cross section order) processes should be taken into account not only for detecting, but also for correctly identifying these phenomena (e.g. an electron-positron pair obtained from a final-state photon). Electrons, muons, and other charged particles such as $\pi^{\pm}$ can electromagnetically interact with atoms, ionising them and releasing secondary electrons that can be collected later by an electric field to produce a signal. Radiation energy loss due to Bremsstrahlung affects electrons more than muons due to their mass difference.

Electromagnetic cascades or showers happen when electrons or photons interact with a sturdy material: these are a sequence of interactions that progressively generate more electrons and photons through pair production, Compton scattering or Bremsstrahlung, with subsequent reduced energy. The shower, if the block of material is large enough, eventually reaches a point where no additional electrons or photons are generated and the energy of the remnants is absorbed by the material through ionisation and excitation. A relatively similar phenomenon happens with hadrons (called analogously a hadronic cascade or shower), but with several differences. Both showers are usually exploited by calorimeters to measure a particle's energy. A more deep review on particle-interaction matter can be seen in [31, 108].

2.2.1 The coordinate system

The common euclidean coordinate system $((x, y, z) \in \mathbf{R}^3)$ is scarcely used in accelerator physics. The usual z axis is fixed through the pipe, with the positive semiaxis pointing to the Jura mountains, x points towards the centre of the LHC's circumference and y towards the sky, orthogonal to the collider's plane. Usually, only z is used, and the xy plane is parameterised by $\varphi \in [-\pi, +\pi]$ and $\theta \in [-\pi/2, \pi/2]$ (φ describing angles between the x and y axis and θ between the z and x ones). Being the origin of coordinates located in the collision point, any location in space can be defined with (r, θ, z), with r the distance from the origin to the location. However, if we are interested in the orientation of particles that come from the collision (for example, when we determine the four momentum of one final state particle of a process), we can precise that using (η, φ), where $\eta := -\ln(\tan \theta)$ is called **pseudorapidity**. It can be proven that this value for the direction of one particle can be written as

$$\eta := \frac{1}{2} \ln \frac{|\vec{p}| + p_z}{|\vec{p}| - p_z} \simeq \frac{1}{2} \ln \frac{E + p_z}{E - p_z} =: y, \tag{2.5}$$

where $\vec{p}$ is the momentum of the particle, p_z the absolute value of the z component of it, E the energy and y the **rapidity**. The pseudorapidity and rapidity are approximately equal as shown in that expression if $m \ll p$, a limit usually accepted in the conditions of experimental HEP. Differences in the η-φ space are commonly used in the field and can appear defined as $\Delta R := \sqrt{(\varphi_1 - \varphi_2)^2 + (\eta_1 - \eta 2)^2}$.

2.2.2 Tracker

Beginning from its core, the first component of the CMS detector is the tracker, whose main mission is to obtain the momentum of charged particles that cross it by measuring the curvature of their trajectories. It also has the secondary mission of locating the interaction point of the interest process, more usually called "primary vertex". It is mainly made of silicon pixels, which offers granularity yielding a precision of $10\,\mu$m. It is cooled down using a gas system to $-20°$C to reduce the aging and the effect of the radiation, as well as the heat due to the high number of connections. The pixels work as 65 millions of receptors that whenever a charged particle crosses them collect the electrons of the ionised silicon atoms and transform them into a signal. A similar functioning happens for the outer part of it, made of a barrel of silicon strips.

2.2.3 Electromagnetic calorimeter

The next subdetector is the electromagnetic calorimeter (ECAL), responsible for measuring the energy of the electrons and photons. It is made of a solid scintillator, lead tungstate, that is disposed in form of crystals, giving the advantage of celerity, as scintillation is a quick and known process. Whenever an electron or a photon enters the ECAL, it will soon interact and

trigger an electromagnetic shower (a sequence of processes due to the interaction of an electron or photon with matter) that will produce scintillation, a process in which photons are released as a consequence of the interaction of the incident particle with matter. Photons are then collected and its signal enhanced in photomultipliers that also transform it into an electric signal. There are of the order of 80000 of these crystals that rise the weight of the electromagnetic calorimeter to the 100 tn. A brief mention deserves the so called preshower of the ECAL, whose aim is to provide a better differentiation in trajectories for particles like neutral pions that might have a high velocity in the z direction after being produced and thus their most probable decay, two photons, too: these two photons would be usually interpreted as one, in these cases. The preshower, located in the endcaps, has a higher granularity that grants a better capability of identifying both of them. It is made of lead and silicon.

2.2.4 Hadronic calorimeter

After the ECAL, the hadronic calorimeter (HCAL) measures the energies of hadrons, i.e., particles made up of quarks, such as kaons or pions. It is a sampling calorimeter, made of successive layers of an absorbent material, and a fluorescent scintillator. The procedure for collecting the light of each scintillator is the same as in the tracker or the ECAL, but now the energy of different layers of scintillating material (that is said to form a "tower") is collected using optical fibres, and then the energy of the hadron is said to be the sum of all the light that the hadronic shower (analogous as the electronic, but more complicated and with more type of processes) produces whenever a hadron enters the HCAL. It is hermetic, in order to not let any kind of particle (except muons and obviously neutrinos) that might have not been stopped yet to go beyond it. This will ensure later that we can obtain information from particles that we cannot directly detect, as will be explained in the next chapter.

2.2.5 Superconductor solenoid

Between the HCAL and the muon system the superconducting solenoid provides a large magnetic field of 4.2 T. It is actually the largest superconducting magnet ever built, with 12500 tn of weight, and its function is essential. Thanks to it, the momentum (through curvature) of charged particles can be obtained, and we can also differentiate between positive and negative charged particles. To guide the field lines outside the solenoid, though still inside the detector, a structure scaffold-like called the return yoke, has been constructed through all the muon system. It is made of iron and helps define the magnetic field lines.

2.2.6 Muon system

Finally, the muon system has been constructed in between the return yoke structure. It is the most outer subdetector due to the fact that muons are expected to be able to exit the hadron

calorimeter, due to their comparative long mean lifetime and the speeds at which they will surge from the collision point. Its main function is to identify the particles as effectively muons, and to measure their momenta by tracking their trajectories, an effort for which their information is crosschecked with the one from the tracker, allowing usually to obtain a good muon identification. There are three main elements that conform the muon system: the drift tubes (DT), the cathode strip chambers (CSC), and resistive plate chambers (RPC). Cross sections of CMS highlighting the components of the muon system can be seen in Fig. 2.9, both transversal (in the barrel) and longitudinal. The latter also shows the new muon detectors installed in the context of the upgrades (described in Sect. 2.4) of the LHC and CMS: the gas electron multipliers (GEM), the improved RPC (iRPC) and the ME0 detector, based on GEM technology.
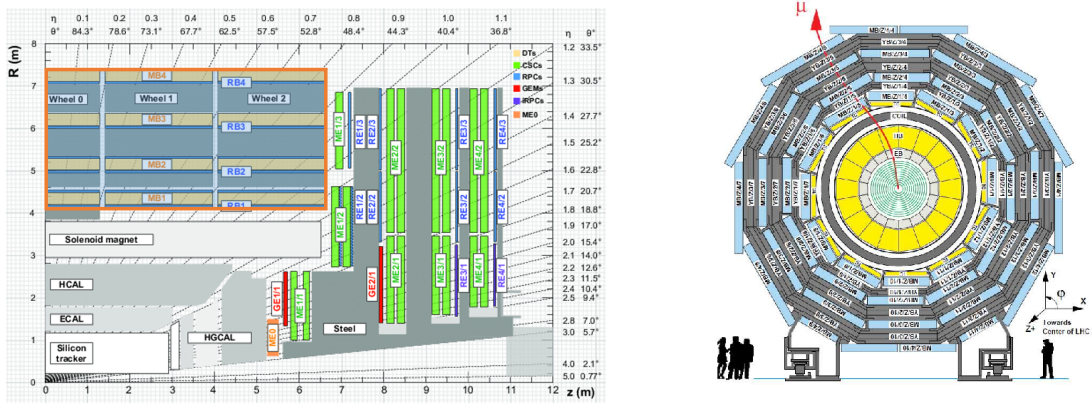
Figure 2.9: CMS' longitudinal[109] (a quarter) and transversal [110] (in the barrel) cross sections highlighting the muon system.

The most important (also for this thesis) and numerous are the DTs. They are located in the barrel and dispose in several groups (chambers) through the return yoke structure. A DT is a cavity filled with gas in which an electric field is present between an anode (a wire through all the tube) and a cathode, as seen in Fig. 2.10. Whenever a muon crosses the gas, it will ionise some gas atoms and thus generate electrons that will drift to the cathode, triggering an electronic signal that is later recognised by dedicated hardware.

One DT chamber is composed of multiple individual DT (cells) disposed in layers (L), usually with ≈ 50 in each. Four stacked layers conform a superlayer (SL). One DT chamber is a set of three SL organised as seen in Fig. 2.11: the externals SLs are disposed to differentiate the φ position of an incoming particle, while the internal is able to assign θ values. A spacer (a honeycomb structure) is included in the middle of one external SL and the θ one.

Cathode strip chambers, unlike DTs, are used mostly in the endcaps. They are chambers also filled of gas, and with a similar working method as the DTs, but their anodes and cathodes are disposed orthogonally, allowing to measure both dimensions in only one CSC, which is an advantage considering that in the endcaps there is no return yoke and the magnetic field is
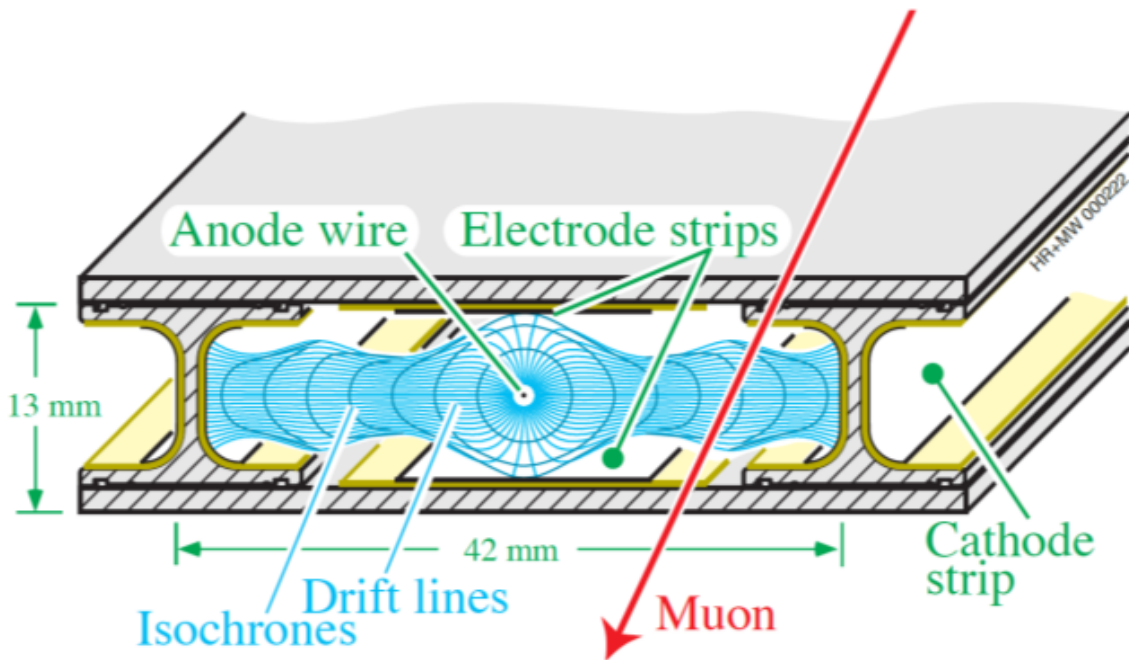
Figure 2.10: [109] Diagram of the transversal plane of one drift tube (or cell), and part of its left and right neighbours.

less uniform, thus identifying particles with higher precision is more necessary. The third main component of the muon chambers are the RPCs, whose working method is also similar to the other two, but far faster. They are distributed in the barrel and endcaps and their information is used for the trigger procedures.

2.2.7 Human and computing resources

The data collected, as we will see later, needs to be stored before it can be used for physics analyses. To do so, there are many computing facilities at CERN, some of them in CMS' site, that do this initial processing and storage. Afterwards, a large amount of computing resources is needed to do all the processing and simulations required to obtain the scientific articles that are later submitted to the journal. These resources are instantiated in the form of large supercomputing centres, and are structured in tiers. The tier zero (T0) is the data processing centre at CERN tier 1 (T1) comprises data centres that have a good, dedicated connection with CERN and share the responsibility with T0 of storing the raw data. Tier two (T2) equipments are used as computation facilities mainly, to be used also by collaborators all around the world for analysing their work. Finally, tier 3 (T3) are small computing resources that are usually located in the home institution of collaborators, and are the place where the final data processing takes place. This whole organisation of data centres is called the Worldwide LHC Computing GRID (WLCG) [111, 112].
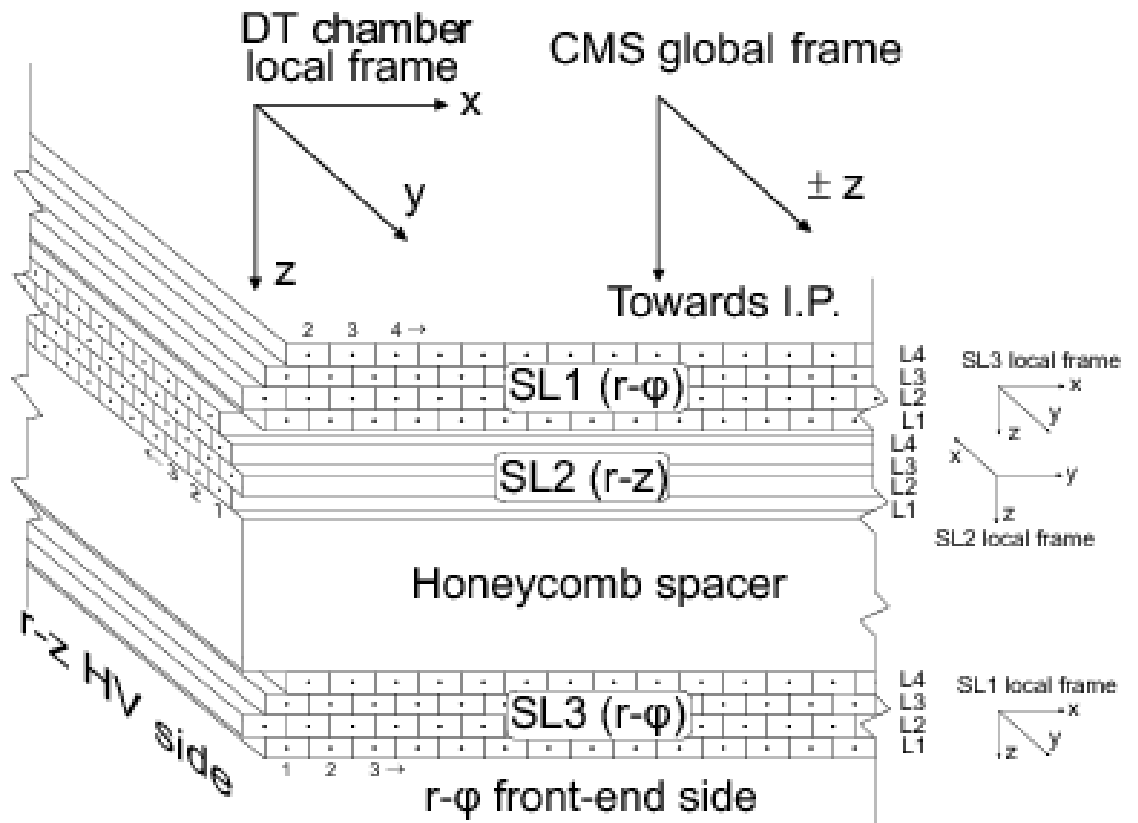
Figure 2.11: [109] Diagram of one DT chamber. Different local and general detector reference frames are also shown.

2.3 Reconstruction

The process of recovering the collision information (mainly, final state particles' four-momenta) from the signals and the information we have collected from all CMS subdetectors (the "raw" data) starts when we have the latter safely stored. Unlike in the trigger system, we do not have any time constrain that forces us to not aim to optimal algorithms for this effort that we name "reconstruction". As mentioned before, this step is applied not only to data but also to the MC simulations we generate, once we simulate the travel of the particles through the detector and the energy deposition and signals that they might leave in the electronics. The reconstruction consists on a sequence of algorithms that begin from the most relevant observables and build up the rest of the event information on them, relying (sometimes) in the information reconstructed by previous stages.

2.3.1 The Particle Flow algorithm

The main CMS algorithm that provides a set of object candidates to be used later for identification is called Particle Flow [113] (PF): a set of steps that progressively reconstructs the kinematic and geometric information from the event. There are two stages in this algorithm: the reconstruction of the particle-flow elements, and then the identification and reconstruction of particles from them. In the following sections we offer a combined summary of these two steps, that can be checked in detail in the provided reference.

2.3.2 Tracks, vertices, and clusters

The first elements that are reconstructed correspond to charged-particle trajectories or tracks. To do so, an algorithm based on a Kalman Filter [114] is used to, iteratively, derive the underlying particles' information from the tracker information. It starts from an initial seed provided by the information of two or three hits on the tracker [115], then this information is used to find more suitable signals (or "hits") in other tracker layers. This process is repeated for up to ten iterations, in increasing complexity of the type of tracks to be fitted, as well as the origin of the seeds.

This provides for each trajectory, a set of hits with an associated estimation of the track parameters. However, they might be biased due to constraints during the seeding step. To avoid bias, the tracks are refitted to extract the final parameters with another Kalman Filter applied sequentially, from the inner hits to those far from the interaction point. The process is combined with a smoothing based on another Kalman Filter applied backwards, from the outermost hits to those closer to the interaction point, and the average of the parameters of both filters are taken. The extrapolation of the next hit location between layers is improved by using a Runge-Kutta propagator that solves first-order differential equations taking into account a mapping of CMS' magnetic field with a $< 0.01\%$ precision. This propagator is used both in the filtering and smoothing stages, and is more relevant in inclinations larger than $|\eta| = 1.0$, due to magnetic

field inhomogeneities. After this, the tracks are pruned from outlier hits, as well as other quality criteria, and the final parameters are obtained [116].

The obtained tracks can be used to derive interaction vertices by extrapolating the reconstructed trajectories to the beam axis. They are ordered by the quadratic sum of the transverse momentum of the tracks associated to each vertex, i.e. $\sum p_{\mathrm{T}}^2$, and we consider the one associated to the largest sum as the **primary vertex**, whereas the rest are considered as **pileup interactions**.

Apart from tracks, the other main element used to later reconstruct the information from the particles in any event are the "clusters" from the electronic and hadronic calorimeters. The clusters are groups of energy depositions, with its energy and direction associated. They are used in the reconstruction and identification processes of electrons, photons, and hadrons (both neutral and charged). To reconstruct clusters, seeds are derivated from cells (elements of the calorimeters that measure the deposited energy) with larger energy than their neighbouring and larger than a set threshold. Then, topological clusters are built by grouping the seed with touching energetic cells. Finally, a grouping algorithm based on a Gaussian-mixture model is used to obtain the position and energy of the clusters.

2.3.3 Muons

The process to reconstruct and identify these particles depends not only on PF information, but also on the muon system, described in 2.2.6. This subdetector allows to efficiently identify muons, thanks partly to the absortion of other particles by the electronic and hadronic calorimeters (designed so that only muons and neutrinos fly through them). Then, the muon information can be reconstructed using the muon system information, and most of the times, also the tracker's trajectories. Depending on this process, the final muon particles from reconstruction can be classified as follows.

Standalone muon: candidate that is reconstructed from information of the muon system only, i.e. DT, CSC and RPC subdetectors, usually requiring compatibility between "segments" (a track candidate from one muon subdetector, i.e. DTs, CSCs, or RPCs) in more than one layer of the system (e.g. there are at least four layers of muon DTs in the barrel of CMS).

Tracker muon: candidate that is reconstructed from information of one trajectory in the tracker, that (fulfilling some conditions on energy) is matched with a segment in the muon system.

Global muon: standalone-muon candidate whose track is matched to another in the tracker (called "inner" track) if the parameters of both are compatible. In that case, the global-muon parameters are recalculated with the hits of both tracks.

Thanks to the well-designed muon system, almost all ($\approx 99\%$) muons are reconstructed as global or tracker muons, and usually as both. After these candidates are obtained, requirements are imposed to obtain the final set of PF muons, that can be of different types. Isolated global PF

muons are identified by demanding that the transverse momentum (p_T) and energy of the tracks and energy deposits in a cone of $\Delta R < 0.3$ (in the $\eta - \varphi$ plane) is less than 10% of the muon p_T. This allows to efficiently reject hadrons whose hadronic decay cascade in the HCAL might slightly cross to the muon system (commonly called as "punch-through"). Muons that do not pass the isolation requirements, and are expected to belong to a jet of particles, must fulfill other criteria to avoid the same misreconstruction with hadrons. The muon momentum is that of the tracker's information if its transverse momentum is less than 200 GeV: above this, comparisons are done with the momentum extracted with the information from the muon system. Reconstruction efficiencies of muon trajectories in the tracker can be consulted in [116] as a function of p_T and η, and are roughly 100% for central muons of $p_T > 1$ GeV. For segment reconstruction in the barrel DT subdetectors, efficiencies are also essentially 100% or $> 99\%$ as shown in [117] with data from 2018 collected from proton-proton collisions at $\sqrt{s} = 13$ TeV.

2.3.4 Electrons

Electron reconstruction and identification depend, as in the case of muons, on whether we are treating jet-belonging particles, or well-isolated ones. For the latter targets, electron reconstruction begins in the ECAL, with clusters with at least 4 GeV of transverse energy. Their position and energy is used to extrapolate potential hits location in the tracker, taking into account that the particle is either a positron or an electron. However, the energetic electrons that are produced tend to irradiate photons due to Bremsstrahlung in their travel through the tracker. Thus, to correctly reconstruct the electron energy, all the ECAL clusters that can be linked to the electron itself as well as all its irradiated photons are grouped in a "supercluster", with a small η and big φ span.

For non-isolated electrons, this ECAL-based procedure can lead to high inefficiencies, and thus another reconstruction is done, initiated in this case by tracker-reconstructed tracks. For electrons that did not radiate in the tracker, their reconstruction efficiency can be as high as for muons, whereas for those that might radiate the looser requirements in the track reconstruction are able to recover them. The tracks are later extrapolated to the ECAL and linked to suitable clusters.

The track seeds from both approaches are then re-fitted for electrons with another algorithm (a Gaussian-sum filter, or GSF) that provides resilience to photon irradiation and thus better fitted parameters. The combination of both approaches for the seeds, and the later GSF fit, allow to reconstruct electrons with even 2 GeV of p_T, although the ECAL-only method provides a smaller misidentification rate with charged hadrons (mostly pions). Electron candidates are selected from both approaches' seeds if an ECAL supercluster is linked to at least a GSF track and not more than than three of them. Other requirements are imposed upon ECAL-based electron candidates to avoid misidentification, such as demanding a maximum in the energy depositions in the HCAL associated with ECAL's clusters, as well as a machine-learning based algorithm. Overall efficiencies of electron reconstruction, as a function of their reconstructed η and p_T, can

be seen in Fig. 2.12 using data taken in proton-proton collisions in CMS from 2017. They have been estimated with the tag-and-probe method, described in e.g. [118, 119].
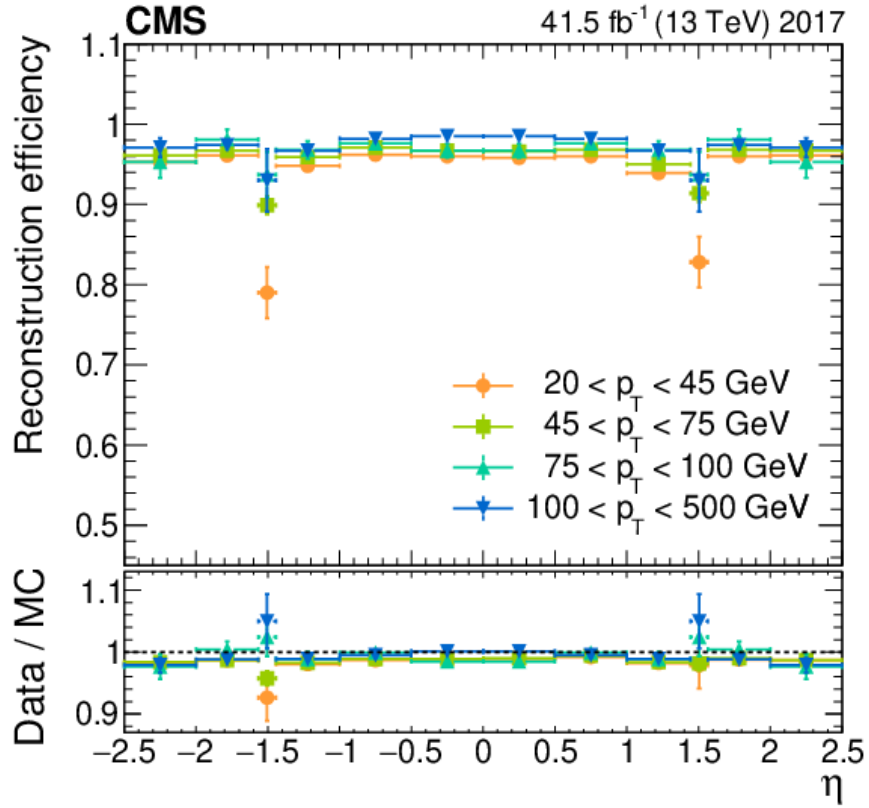
Figure 2.12: [119] Electron reconstruction efficiencies as a function of their η and p_T estimated with proton-proton collision data at $\sqrt{s} = 13$ TeV taken by CMS. The vertical bars show the total (statistical and systematic) uncertainty.

2.3.5 Photons

Reconstruction of photons, due to its relationship with electrons, is done at the same time as them. Isolated photons are identified as such from ECAL superclusters with $E_\mathrm{T} > 10$ GeV and without any link to a GSF track. Other requirements are imposed to linked energy deposits in the HCAL, in a similar way to electron requirements.

Non-isolated photons can appear in CMS' events coming from e.g. π^0 decays. Any ECAL cluster (not already used to identify electrons, or isolated photons) inside the tracker acceptance is assumed to be a photon, as studies show neutral hadrons (that might leave some of their energy in the ECAL too) are expected to leave a very small fraction of the energy in the ECAL. Out of the tracker acceptance, charged and neutral hadrons cannot be differentiated and the energy fraction they leave in the ECAL is considered as part of the same hadronic shower, if they can

be linked with another HCAL cluster. If not, they are assumed to be a photon. Other criteria imposed to assign the final energy to the hadron and photon candidates.

2.3.6 Hadrons

Hadrons are reconstructed after the identification of muons, electrons and well-isolated photons: the remaining particles to be reconstructed and identified are charged and neutral hadrons, non-isolated photons (already treated in the previous subsection) and, less often, also muons decayed from charged hadrons.

Inside the track acceptance, as for the case of photons, HCAL clusters not linked to any track are reconstructed as a neutral hadron. Outside the tracker reach, HCAL clusters are reconstructed as a hadron with the information associated of any ECAL clusters that can be matched with it.

2.3.7 Jet clustering and identification

As mentioned in previous sections, final-state produced quarks undergo the process of hadronisation, creating after the collisions a collimated group of particles (or jet) because of the nature of the strong nuclear interaction. There are many processes of interest (such as $t\bar{t}$) where final-state quarks or their information are essential for many reasons, and thus a way of recover the original final-state four-momenta of quarks was developed. To do so, we reconstruct the jets originating from quarks through a clustering, anti-k_{T} [120, 121], that uses (for the set of jets we are using in our analysis) all particles reconstructed by PF. These jets or "PF jets" have at least 15 GeV of p_{T} and, on average, 65% of its energy is carried by charged hadrons, 25% by photons and 10% by neutral hadrons.

2.3.7.1 Jet tagging

Both gluons and any of the six quarks (and their antiparticles), except the top, can generate a hadronic shower and thus a jet after a collision. Differences among quarks (e.g. their mass) and with gluons force that the features of the jets originated from them are also dissimilar. Experimentally, it is of interest to be able to identify, somehow, from which flavour a jet originated, as it allows us to pinpoint the final state particles. A motivation for this can be derived from the layout of the CKM matrix: in top-related analyses, this quark will almost always decay to a b quark and a W boson and thus if we can detect events with jets that originated from a b quark we might have a handle on a top quark production process.

Because of the differences among quarks, algorithms that link (or tag) jets to a precise quark flavour are usually developed for b quark-originated jets. Its large mass (the largest, except from the top quark) forces that the hadron that is formed shortly after its final-state production has a slightly larger mean lifetime, and thus it can travel a bit further away before decaying into other particles whose track can be later reconstructed by CMS and its subdetectors (e.g. pions, muons).

This creates another point of interaction, near the actual collision, that can be reconstructed thanks to CMS' tracker precision: we call it a secondary vertex. Depending on the separation between the secondary and primary vertices (the impact parameter d) and other features from the event (e.g. momenta of the jet components), etc., an algorithm can be constructed to identify jets that come from a b quark. This procedure is called **b-tagging**. There are also algorithms that attempt to tag jets that originate from a c quark, light-flavour quarks[1] or gluons.

Nowadays, most algorithms that do jet tagging (particularly, b-tagging) are based on machine learning (ML) techniques that provide a way to take advantage of correlations among many different variables in complex ways, sometimes exploiting very hidden patterns from data used. For analyses in this thesis, we use the `DeepJet` [122] algorithm, that exploits roughly 650 variables into a deep neural network that consists of various convolutional and long short-term memory (LSTM) layers. It is a multi-classifier, with six output nodes to tag a jet as coming from a bottom (with three different nodes depending on how the jet originated), charm, light-flavoured quark, or a gluon. Applying the model to the jets in the events, values (probabilities) can be extracted from the output nodes and from them, estimators for e.g. c-tagging can be constructed as

$$C := \frac{P(c)}{P(c) + P(uds) + P(g)}, \tag{2.6}$$

where $P(p)$ is the output value of the node for the p particle. For b-tagging, the sum of the probabilities of the its three final nodes is used. A review of the performance of `DeepJet` can be seen in Fig. 2.13.

2.3.8 Taus

Unlike their other (significantly massive) leptonic colleagues, electrons and muons, taus are practically unable to reach the tracker subdetector because of their very small lifetime, with is $\tau(\tau) = (2.903\pm0.005)\cdot10^{-13}$ s, compared to that of the muon $\tau(\mu) = (2.1969811\pm0.0000022)\cdot10^{-6}$ s [31]. Thus, taus decay electroweakly before reaching the tracker through two different set of particles. Either a "leptonic" decay, into either electrons or muons plus associated neutrinos, or a "hadronic" one. This last production mode is allowed for taus (and not electrons or muons), because of the comparative large mass of this particle ((1.77686 ± 0.00012) GeV) [31], thus providing the enough four-momentum to an intermediate W^* boson to decay later into quarks that afterwards undergo hadronisation.

Taus that decay leptonically are almost impossible to identify, because the other particles from the decay are neutrinos (that are not detected by CMS). Thus, we consider electron and muons from tau leptonic decays as final-state leptons from the interactions (or "prompt" leptons) in

---

[1]Apart from the top quark, the bottom and charm are considered **heavy-flavour** quarks due to their masses of $m_b = 4.18^{+0.03}_{-0.02}$ GeV and $m_c = 1.27 \pm 0.02$ GeV, various orders of magnitude larger than those of the up, down and strange quarks ($m_u = 2.16^{+0.49}_{-0.26}$ MeV, $m_d = 4.67^{+0.48}_{-0.17}$ MeV, $m_s = 93.4^{+8.6}_{-3.4}$ MeV), dubbed accordingly **light-flavour** quarks [31].
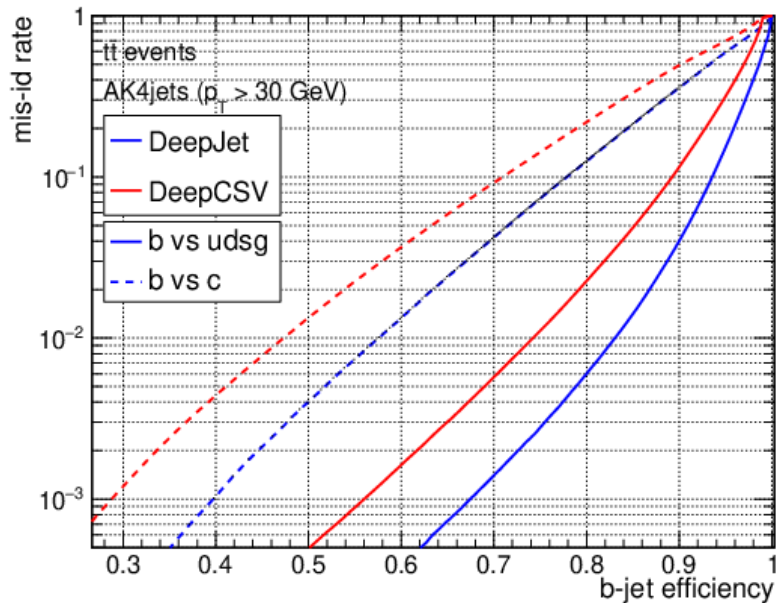
Figure 2.13: [122] Misidentification rate with respect to b-jet tagging efficiency for the `DeepJet` algorithm and also `DeepCSV` [123] algorithms. They are evaluated using simulated $t\bar{t}$ events and jets clustered with the anti-k_T algorithm with $\Delta R < 0.4$ and at least $p_T < 30$ GeV. Continuous lines show the performance of b quark-originated jet tagging against light-flavoured quark and gluon-induced jets, while dashed lines show b quark-originated tagging vs c-quark jets.

our analysis. However, hadronic tau decays, that yield jets of particles in the detector, can be identified taking advantage of the properties of the jets that differentiate them from others in the detector, such as the hadrons that they contain or its collimation. This algorithm reconstructs then the four-momentum of the tau and provides a set of "hadronic tau" particles.[2] Electrons and muons coming from tau leptonic decays are considered in the analysis of this document as any other leptons that might come from a W or Z boson decay.

2.3.9 Missing transverse momentum

As neutrinos only interact through the weak interaction (and gravity, but as we mentioned before we are neglecting its effect in the context of high energetic particle collisions), they hardly ever leave any signal in any CMS subdetector. However, they are present in many final states of interesting processes, as well as they can participate in intermediate steps of them. A way to probe the neutrino final state information is using the conservation of four-momenta in interactions, and thus deriving what is commonly called the missing transverse momentum p_T^{miss}. This variable

---

[2]A consequence of the behaviour of taus in high energy colliders, as well as the neutrino non-observation is that we commonly call "lepton" only to electrons and muons, as it might be used in the rest of this document, emphasising those situations where taus and neutrinos are included as "leptons".

allows to have a proxy, in the transverse plane, to all non-observed objects in an event (such as neutrinos). p_T^{miss} is defined including the PF energy correction to jets as

$$p_T^{miss} = - \sum_{i=1}^{N_{particles}} \overrightarrow{p}_{T,i} - \sum_{j=1}^{N_{PF\ jets}} \left(\overrightarrow{p}_{T,j}^{corr.} - \overrightarrow{p}_{T,j} \right), \qquad (2.7)$$

where $N_{particles}$ is the total number of PF reconstructed particles, $N_{PF\ jets}$ is the total number of reconstructed PF jets and $\overrightarrow{p}_{T,j}^{corr.}$ is the PF corrected p_T of the j-th jet.

Apart from serving as a potential source of neutrinos information, p_T^{miss} also serves to quickly detect those events where the reconstruction of a particle was badly done, or where the presence of collision background, such as cosmic rays, interferes with the interaction of interest. Thus, a series of **filters** are developed based on the p_T^{miss} observable: either on its value itself, or its variation with the removal of one reconstructed particle (e.g. a cosmic muon, or a high-p_T one). In Fig. 2.14 it is shown the effect these filters have in the distribution of p_T^{miss} and also in the φ of the reconstructed jet with largest p_T.
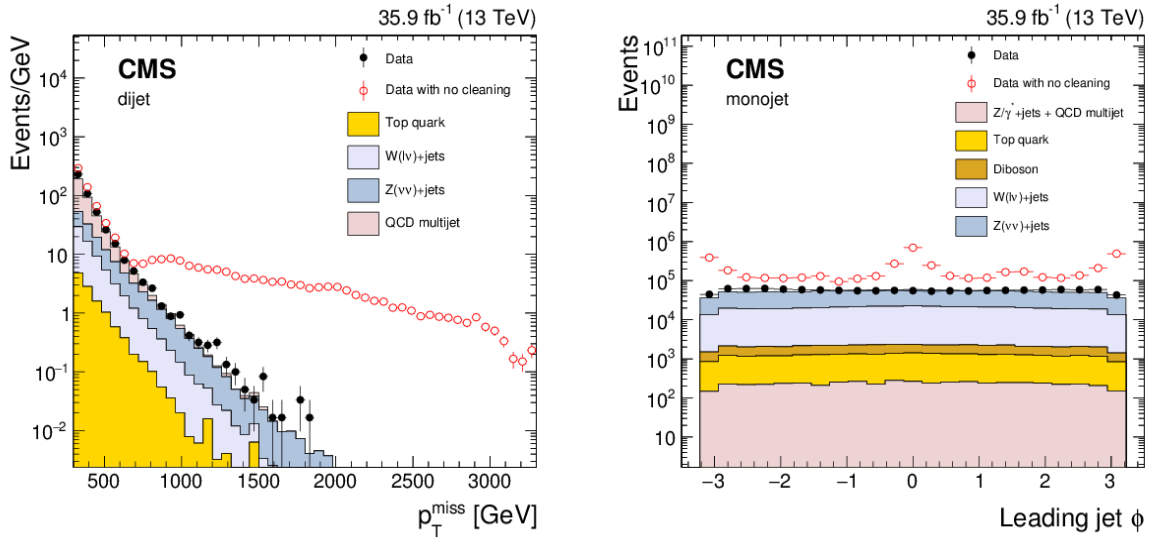
Figure 2.14: [124] Histograms of the p_T^{miss} (left) and reconstructed jet φ with largest p_T (right). Data with p_T^{miss} filters applied is shown as black points, whereas events with no cleaning use hollow red ones. MC simulations from various processes are shown for comparison. The left plot has an event selection aimed to have two jets, whereas the right plot targets events with only one: the detail of the selections can be found in [124].

2.4 The high-luminosity upgrade

As a continuation of the work done and predicted for the current LHC and its experiments (that should last until 2029), an upgrade to the accelerator and its experiments has been designed. This

intervention aims to increase the working luminosity of the collisions at the LHC, and to adapt the detectors to withstand the new conditions, as well as general improvements. Consequently, the project is called high-luminosity LHC (HL-LHC), and should in essence provide an increase in the total data able to be collected. Since its beginning in 2011, the endeavour has been done in the design and preparations side, that ended with the final version of the HL-LHC technical design report (TDR) [125], that covered the actions to be taken only on the collider. In 2018, the civil-engineering works started [126]: the core elements are the use of more powerful magnets and optics to squeeze even more the beams thus enhancing luminosity. An overview of the planning for the short term and long term of the LHC and its experiments can be seen in Fig. 2.15.
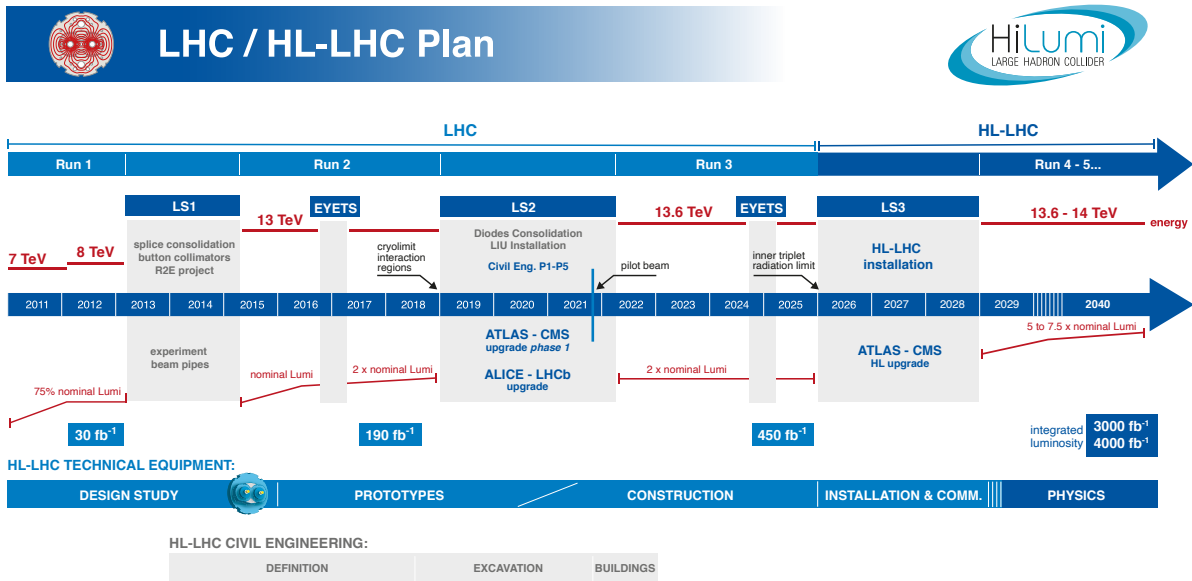
Figure 2.15: [127] Outline of the LHC and HL-LHC working and upgrade plan.

The experiments worked on their preparations and potential updates also, and released TDRs accordingly. For the case of CMS, various documents have been released for different elements: the beam radiation instruments and luminometers [128], tracker [129], barrel calorimeters [130], endcap calorimeters [131], muon system [132], a new timing detector [133], the L1 trigger system [134], and the data acquisition and high level trigger [135]. The main idea of CMS' upgrade is to prepare the detector for the high-luminosity conditions, improving the experiment in the process. A higher instantaneous luminosity has various implications. The first of them is an increase in PU. These extra simultaneous collisions will leave signals in the subdetectors, thus enlarging the information per event in all data-collecting channels. A larger $\mathcal{L}$ also means additional irradiation of the detector itself and thus more wear for the materials. The following subsections contain a brief description to the study of the ageing effects of the DT subdetectors

on the reconstruction efficiencies.

2.4.1 Ageing of the DT subdetector and effects on muon reconstruction

Studies on the longevity of the DT chambers of CMS have been undergoing to estimate how it will be affected by the Phase 2 upgrade: an early public release of them was made in [109], although internal development continued and yielded the results of [136]. This research was done using a spare DT chamber that was irradiated using the GIF++ facility at CERN to a dose equivalent to two times the expected integrated luminosity for the entire HL-LHC ($2 \cdot 3000$ fb^{-1}). A radioactive source of ^{137}Cs was used to do so. Although the entire chamber faced radiation, only some of the cells had an active voltage differential between anode (wire) and cathode (strip): the rest had voltage applied, but with negligible gain. Thus, these cells could be used as control ones.

Soon after the irradiation started variations were seen in different figures of merit used, such as the normalised current recorded in the anode, or the hit efficiency (defined by signals in single cell wires). Examples of such can be seen in Fig. 2.16, where it can be seen that the non-active DT cells showed no change or a small one when checking the anode current and the hit efficiency (that of generating a signal when a charged particle crosses one DT cell).
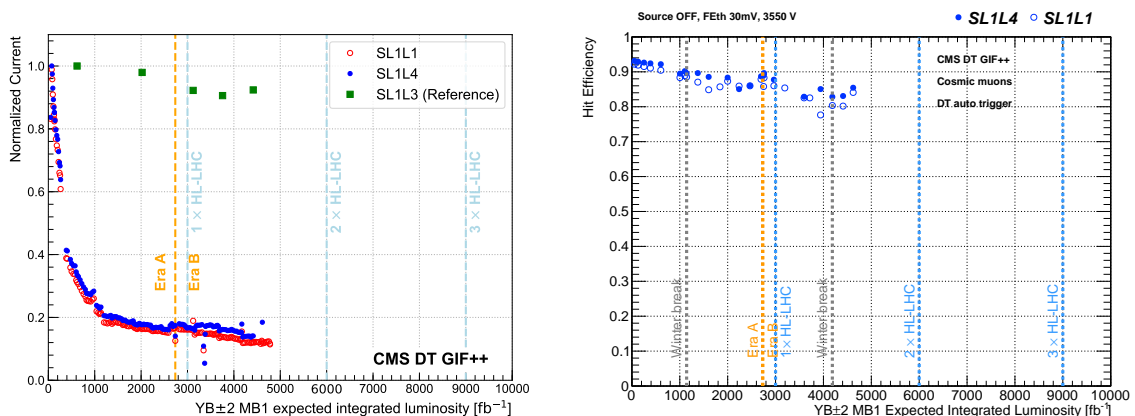
Figure 2.16: [136] Variation of the normalised current registered at the anode from the irradiated DT chamber (left) and the hit efficiency measured with cosmic muons with events triggered by tracks in SLs different to that used to calculate the efficiency (right) against the expected integrated luminosity that would been seen in the MB1 chamber of the wheels ±2.

Although ageing effects could be seen, the impact for physics can be estimated from the muon reconstruction efficiency (that of reconstructing a muon from the signals it left in the subdetectors). The main reason is that the whole muon system depends on several DT chambers, as well as other detectors, and the consequences on muon reconstruction can be directly propagated later to analyses done. Subsequent internal studies incorporated the variation of hit efficiency depending on θ, not previously considered.

To estimate the ageing effect various deterioration scenarios were built, where hit efficiencies
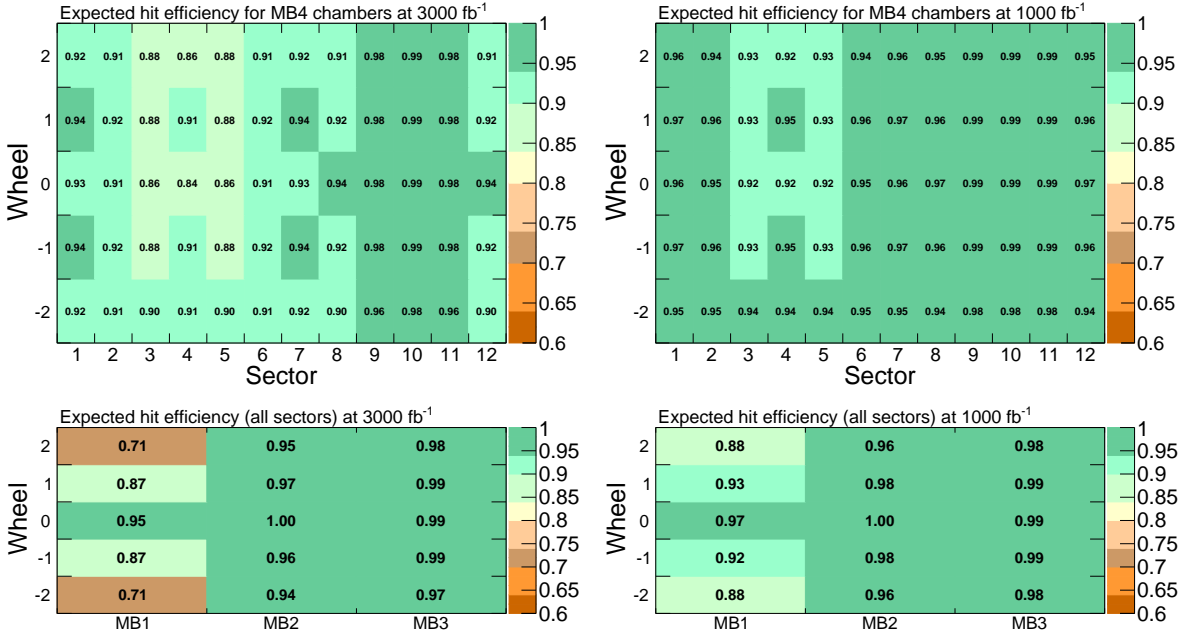
Figure 2.17: [136] Hit efficiency corresponding to ageing scenarios estimated for the DTs when CMS collected 3000 fb^{-1} of integrated luminosity (left) and 1000 fb^{-1} (right).

reduction were assigned to each DT chamber, depending on its location in the detector. Two scenarios were built, one equivalent to 1000 fb^{-1} of integrated luminosity and other to 3000 fb^{-1}, that are displayed in Fig. 2.17 [136].

An estimation of the reconstruction efficiency was done by generating a muon MC sample, with two back-to-back muons of $5\,\text{GeV} \leq p_\text{T} \leq 100\,\text{GeV}$. The sample was sequentially created, and the reconstruction was done with different settings to model the predicted ageing effects. To do so, the assigned hit efficiency in the scenario was transformed in a probability, so that all simulated signals on each DT cell had that chance to be recorded or not. One MC sample was entirely generated with nominal settings (i.e. without modelled ageing), while other five used the same generation information but with different ageing configurations. One used the scenario of 3000 fb^{-1} depicted on [109]. Another one used the scenario of 3000 fb^{-1} developed since the previous result and from Fig. 2.17. An additional simulation added, on top of this scenario, modelling of failures in the RPC detectors. The last two simulations were done with the 1000 fb^{-1} ageing modelling from Fig. 2.17, and by also adding to it potential RPC failures. The reconstruction efficiencies, depending on $|\eta|$ and φ, are shown on Fig. 2.18.

The most recent results show that, although the hit efficiencies in the scenarios as the one from Fig. 2.17 can go down up to a $\approx 30\%$ decrement, the reconstruction is not as much affected at all. Overall, the worse expectations for central muons are slightly less of a 2% reduction when comparing with all simulations. If the four most recent (that include potential RPC failures) are observed, the worse reductions can be reduced to a $\approx 1\%$, focused on the overlap region. The
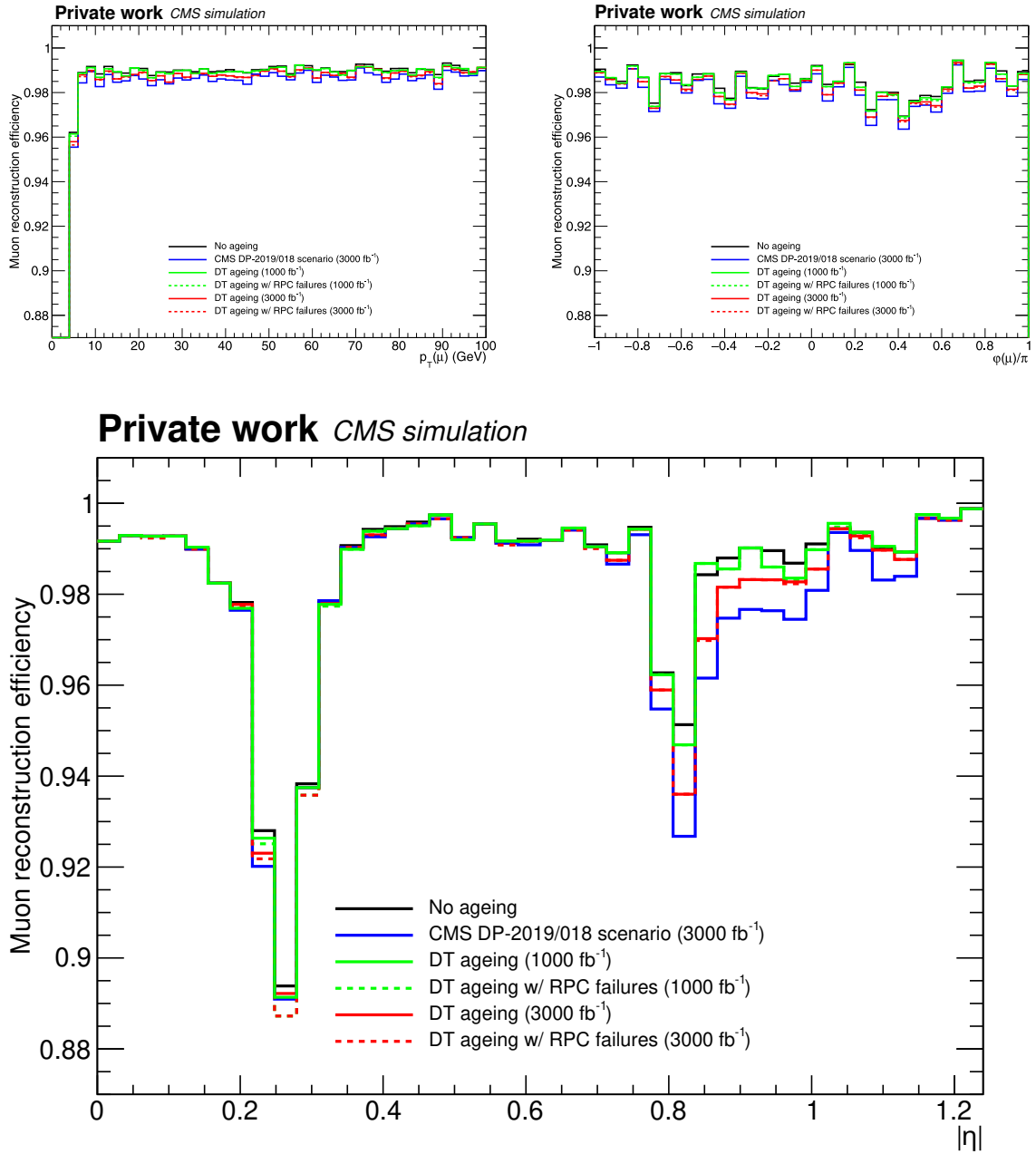
Figure 2.18: Reconstruction efficiency as a function of the p_T of the muon (top left), its φ (top right), and its $|\eta|$ (bottom), calculated with an MC dimuon sample with $5\,\mathrm{GeV} \leq p_\mathrm{T} \leq 100\,\mathrm{GeV}$.

reasons behind no having almost impact in the reconstruction efficiency lie on the redundancy of the entire set of DTs (and RPCs) chambers. The efficiency with the p_T or φ shows no significant variation across the ageing scenarios.

In any case, various mitigation strategies are found to be useful to increase the longevity of the system. Examples of them are modifying the voltage supplied to the cells (e.g. lowering the anode voltage, or modifying the minimum energy threshold used by the electronics to record a signal), adding extra shielding to the detector (to reduce background radiation such as from neutrons), or changing the gas flow across the chambers.

2.5 General methodology of a HEP physical analysis

Collisions that take place in colliders such as the LHC, provide us with raw data. This information is not in an usable format, and we cannot compare it right away with predictions from our theories, usually our aim. One recorded event is usually made of digital information from the particles that originate from collisions in all the experiment' subdetectors. As we are unable to gather and store all the collision information that happen in a collider like the LHC, any datum collected has been preselected by a system, called **trigger**, that quickly (in few nanoseconds) detects an interesting collision.

Thanks to our knowledge of how the particles interact with matter, we are able to derive the physical information of the particles that left those signals (their four-momentum), just after the interaction took place. This process is called **reconstruction** and it provides us with a set of information that allows us to do physical analyses. Particularly, as we mentioned with regards to Eq. (2.2), we can link the counts of collisions with physical properties through the cross sections of processes. When doing physical studies, we focus ourselves in one particular interacting routine, e.g. the Higgs boson decay to two photons: we thus would like to get the data from these collisions, and not the others, so that we can e.g. count how many we have, and compare that number with the expectations. Unfortunately, the interactions that might occur are several, and the data we collect will contain many different production modes. We must then analyse these data, to filter those events, where we think that our process of interest (or **signal**) did not occur. We name as **background** those processes that are not of interest to our study.

There are a lot of properties from the reconstructed final state particles, i.e. from their four-momenta, that we can use to filter the events: their direct momentum, the combined four-momenta properties (e.g. the invariant mass of the sum of two four-momenta) or even the number of an specific type of particles that we found (e.g. number of photons). Before applying these requirements or "cuts" (as they are commonly called), we usually impose restrictions to what we consider a type of particle, so that we are sure (or try) to overcome experimental inefficiencies. For example, to "define" what we call muons in one study, we might require its associated trajectory inside the muon system of CMS to match a large number of signals (or "hits") that must be

correlated with others in the tracker subdetector.

After this **object identification** step is done, we can define in which is at the end the phase space of the final state particles (i.e. their four-momenta) a **signal region** where one could expect to have only signal events, or a very large concentration of them. It is highly likely that we will still have some background processes events, despite our efforts to avoid them: there are two main reasons of this, related with the type of backgrounds we can have. The reducible backgrounds are those that we can suppress by applying tighter requirements on our signal region, because their expected final state can be separated thanks to our experiment detector capabilities, whereas irreducible backgrounds are those that we cannot. Thus, if our process has an irreducible background, we will always have background in our signal region, no matter what. And, if we do not have them, it is highly likely that we still have background processes events, because first of all, our detector is not perfect, and mistakes in the observation of particle properties (in their reconstruction) might occur. And secondly, we might not want to impose very strict requirements into our signal region, as this might imply a very reduced set of events in it (though the signal region might be very pure) and this could thus affect the uncertainty estimation of the measurement we are doing (e.g. the cross section of the process), or maybe we want to analyse data that has some properties such as low energetic (in the context of the collisions we are considering) leptons, and having those in our signal region could imply a larger amount of background events.

Consequently, we will (almost) always have background events in our signal region. Thus, a way of deriving the amount of signal events from the total we have is needed. This **signal extraction** step is done in different ways, but in all of them an estimation of the contributions of background processes is required. To achieve this, we **generate Monte Carlo simulations** with which we obtain pseudorandom-number-derived simulated events, i.e. simulated four-momenta of final state particles of the different processes. This information is not directly useful to us: it would not be fair to compare it with the reconstructed information from data because it has not undergone the same process, with the detector inefficiencies. We therefore simulate how this generated final state particles would travel through the detector, as if they were real, using dedicated particle-matter interaction modelling software. With it, we obtain for each event an identical set of information consisting of signals in each subdetector, corresponding to what the simulated particles would have made, and then we apply the same reconstruction we use for real data to the MC simulations.

This allows us to re-do the entire process we did for data, for the MC: the same object identification and event selection, and then compare data with the predictions (simulations). There might be differences still between data and predictions, and some of them could be due to the different efficiency that the detector, or some techniques might have in data and simulations. To fix these differences, we apply **corrections** to our MC to make a fairer comparison. Other corrections, related with the calorimeter calibrations, are applied also to data. Simulations are also

commonly reweighed, as to avoid statistical fluctuations we usually generate several more events than the ones that could be expected for a specific process. In addition, many times theoretical predictions with a higher precision in perturbation theory for a cross section are available, whereas the generator only achieves lower ones: it is common practice to normalise the MC simulations to the higher precision prediction, yielding a more accurate total normalisation, although the shape effects on distributions still correspond to a lower order in perturbation theory.

With the simulations and data we can now extract the events that correspond to the process of interest from our signal region, always considering that we can model acceptably the background events with the MC. There are various techniques to do this, some of them commonly implying **statistical methods**. The amount of signal events can then be used to extract the signal region, confront their distributions with multiple new models, and thus test theories with data. This however must be done taking into account, on the one hand, the efficiency (ε) of the detector, that is not perfect and prevent us from correctly detecting all events from a particular process. And, in the other hand, the acceptance ($\mathcal{A}$): as our detector is not able to detect all potential phase space coordinates (we do not see the whole 4π sr), we must take into account that there are events that we are loosing that way. Following this, a more experimentally accurate formatting of Eq. 2.2 would be:

$$N = L \cdot \sigma \cdot \varepsilon \cdot \mathcal{A}. \tag{2.8}$$

2.5.1 Implementation

As we mentioned, a powerful computing infrastructure is needed to process and leave ready-for-analysis the data taken at CMS. And also, to generate, simulate the interactions of particles with the detectors, and reconstruct exactly as data the MC.

Physicists that study the data with the simulations need still to do a processing to obtain, in the easiest case, both filtered to the desired signal region. Maybe then a maximum likelihood fit will be done to extract the signal, and afterwards an unfolding with e.g. `TUnfold`. All these steps need also to be adorned with the proper corrections to both data and MC.

All this user-side work needs a proper software framework to be able to process it. The one used in the bulk of this thesis is `CMGTools` [137], done in a mixture of `C++`, `Python 2`, `Bash`, and mainly the library `ROOT` [138]: a very helpful HEP toolkit.

Machine learning models used in this thesis have been trained with `TMVA` [139]: a toolkit prepared to train different algorithms designed to have a good interface with `ROOT`. Maximum likelihood fits are prepared with a `RooStats` [140]-based program [141]. The minimisation step is programmed with the help of the `Minuit` algorithm [142], included in `ROOT`.

3 Trigger, data, and MC simulations

T HE pace of collisions at the LHC is usually high, having collisions each 25 ns (a rate of 40 MHz). The amount of data that the CMS experiment can collect with all its subdetectors per collision is not able to cope with such large collision rate. Thus, we select among all the collisions that take place those that might be the most interesting to us. The whole set of hardware and software needed to detect which collision to store (and which one to not) is called **trigger**. A summary of this system is shown in the first section, including the trigger settings used for the analysis and developments done to the Phase 2 studies for the HL-LHC concerning it. Afterwards, the generation of MC samples is covered as well as the simulations used in the measurements.

3.1 Data triggering

The CMS detector trigger system is divided in two steps: the level one trigger, or L1, and the high-level trigger, or HLT. The L1 trigger, being the first to act, must take quick decisions to not fall behind the LHC due to latency. Thus, it is based in dedicated hardware such as application-specific integrated circuits (ASICs) or field-programmable gate arrays (FPGAs), allowing it to have a buffer of events that can have up to 4 μs of latency while making the decision on them. Each calorimeter and the muon system have their own L1 trigger, where its information (signals, energy depositions) is collected to build "trigger primitives", or candidates of particles. Then, the information from these is combined to do a coarse reconstruction of particles, such as leptons or jets, before a "global" trigger makes the decision whether the event is worth to be kept or not. The decision can be motivated by different requirements, e.g. having one or two energetic jets, or maybe a very energetic muon, thus we sometimes refer to various "triggers". This step reduces the rate of information from 40 Mhz to $\approx$ 100 kHz.

The second stage, the HLT, does not have so large speed requirements as the L1 and thus can run on commercial computing hardware: a cluster located in the same complex as the experiment is responsible of running it. If the L1 trigger makes a positive decision about an event, its information is passed to the HLT, that starts doing a more detailed reconstruction of the detector signals than the L1. This process has conditionals, so that collisions that have passed the L1 stage under a more detailed scrutiny can be rejected if they are of no interest. Roughly $\approx$ 1 kHz of events are stored after the HLT stage, though their processing time can drastically vary

from one to another, as the HLT algorithms that are run, or "paths", depend on each event, and although for some events the decision can be quick, others might spend even seconds in the HLT. The whole set of paths of the HLT that is running alongside with the detector is called the trigger menu. Once the HLT decides that an event must be stored, it is transferred to the T0 storage, at the data centre of CERN.

3.1.1 Trigger configuration used

HLT paths designed to select events with two energetic leptons have been considered for this analysis. These requirements aim to have one electron with $p_T > 12$ GeV and one muon with $p_T > 23$ GeV, or one muon with $p_T > 8$ GeV and one electron with $p_T > 23$ GeV. They are complemented with other trigger requirements that filter events not having (at least) one energetic lepton in the events. These single-lepton triggers aim to have either electrons or muons with cuts in the $24 - 32$ GeV of p_T range. This slightly helps to increase the total efficiency, that is $\approx 98\%$. As mentioned before, roughly $138\,\mathrm{fb}^{-1}$ of data are used in the analysis, coming from the 2016, 2017, and 2018 years (from the Run 2). The luminosity recorded as a function of timeand the pileup of these datasets can be seen in Figs. 2.4 and 2.6.

3.1.2 Ageing effects of the DT subdetectors on the L1 muon trigger

Another aspect of the experiment that suffers the ageing of its components is the trigger. The impact in this system is more relevant than on reconstruction, as a degradation in the trigger impacts the amount of data collected, whereas for reconstruction new algorithms can always be developed, as the raw data is stored.

The HL-LHC will bring new electronic systems to the muon barrel subdetectors, and to exploit their advantages, new algorithms are being developed. Studies to analyse how the expected ageing of the DT impact them have also been done, partly with my collaboration. Two algorithms appear in these comparisons: the analytical method (AM) [143] and the Hough transform-based (HB) [144] algorithms.

The AM algorithm updates the existing Phase 1 L1 algorithm for the DTs, being able to take advantage of the new electronics that allow to have the total direct information from an entire chamber (and more) in one step. The current trigger algorithm is divided in various sequential steps, done by different programs. It aims to get for each trigger primitive (TP), or muon candidate, the same information as the AM or any other algorithm: its global inclination in φ with respect to the interaction vertex, and the slope of the trajectory the muon had in the chamber[1]. First, the bunch and track identifiers (BTIs) use the information from the SL1 and SL2 (separately) to identify both BXs and create segments (sets of aligned hits in all four layers) for each crossing. Afterwards, the track correlator (TRACO) correlate the segments from both

---

[1]It is important to note that, given that the magnetic field of the detector is channeled in the return yoke (a steel structure), muons travel the DT chamber as straight lines.

SLs (if possible). Then, a ghost-reduction and selection of the best candidates per chamber is done by the trigger server (TS). Afterwards, the (local) information from each DT chamber is collected in each sector before passing the information to the total DT track-finder trigger (and then, to the global trigger). An schema of the first steps, corresponding to the DT local trigger, is shown in Fig. 3.1.
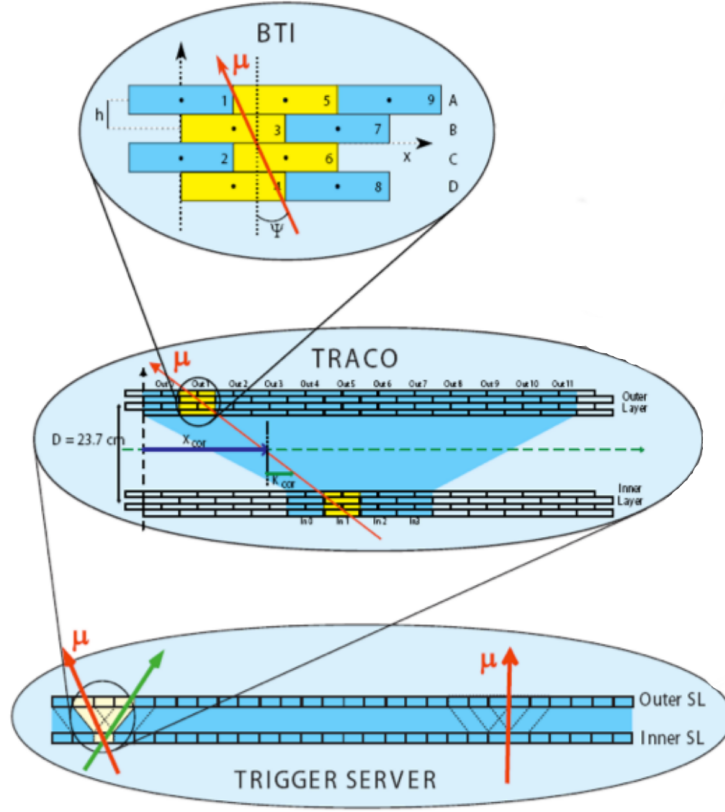
Figure 3.1: [145] Schema of CMS' Phase 1 DT L1 trigger.

The AM algorithm, instead, follows other steps, using as input the time of all hits (taking as reference the start of the LHC orbit, i.e. a moment fixed by the collider as the beginning of one circulation of the beams to the entire accelerator). Here we briefly describe them, as our group worked closely with its development:

1. For each SL1 and SL3, a search of segments of three or four hits is done using a total of 10 cells at the same time (the space needed for all possible physical trajectories). If a segment of three hits is found, the time is extracted directly from it. Concerning the laterality, i.e. the side (left or right, as seen in Fig. 2.10) from which the ionised electron(s) came (and also the particle crossed), all possible physical possibilities are considered. In the case of four hits, the time is calculated for each three-hits subset and the average is taken, and for the lateralities the option with the lowest χ^2 estimator from a fit of the hits location to it is chosen.

2. Then, a correlation is done between groups in both φ SLs if both segments fit in a ± 25 ns time window as well as have physical compatibility. If that is the case, the information from both segments is combined, and a final muon candidate, or TP, is derived. If no correlation is found for one individual segment, it is considered as a final TP also.

The new electronics allow for extra steps to be added, such as including information from the RPCs of the system. This is doable e.g. in an extra step where the final TPs are correlated with hits in the RPCs, to take advantage of its time resolution. Another possibility is to include information from the θ SL (SL2), to enhance its trigger efficiency. Options like these can be use to make more robust the algorithms to ageing effects.

The HB algorithm, instead, opts for a very different approach, based in using a mathematical tool called Hough transform [146, 147]. It was developed originally for HEP, to recognise patterns in a bubble chamber, and has been used later in the field (e.g. [148]), as well as in others (e.g. [149]). The idea is to move the information of hits to a different space, where points in a 2D (x, y) space can be transformed to a (ρ, θ) pair verifying

$$
\begin{aligned}
\rho &= x \cos \theta + y \sin \theta, \\
x &= \rho \cos \theta, \\
y &= \rho \sin \theta,
\end{aligned}
\tag{3.1}
$$

where ρ is the length of a segment between the (x, y) point to the origin, and θ is the angle of such line with respect to the x axis. An illustration of the transform can be seen in Fig. 3.2.
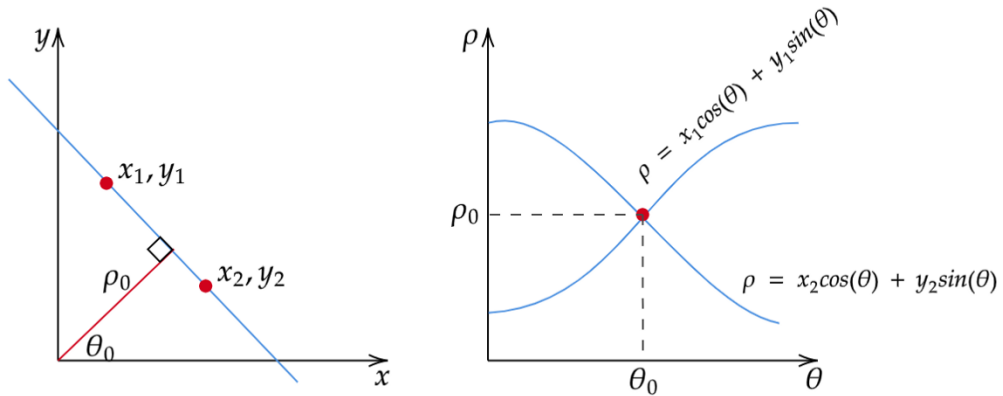
Figure 3.2: [150] Graph showing the direct/point (x, y) space (left) and the dual/line (ρ, θ) space (right) for two points (x_1, y_1) and (x_2, y_2). The points are marked in red in the direct space, and its corresponding lines are shown in the dual as a dependence on θ. The point where they cross in the line space correspond to the (ρ_0, θ_0) values that in the point space define the closest segment to the line that crosses both.

The interest of this transform is that lines in the direct (x, y) space can be linked with points in the dual (ρ, θ), and viceversa. Thus, by unrolling the $\theta \in (0, \pi)$ interval in the first equality of Eq. (3.1), we can draw a line in the dual (ρ, θ) space that corresponds to all the possible (straight) lines that cross the (x,y) point. Thus, finding the $(\rho_{max}, \theta_{max})$ values where the lines of the hits of a segment coincide, we can get a line in the direct space that crosses all the hits (and would describe the trajectory of the particle). Usually, implementations of the Hough transform use binned 2D histograms in the dual space, so the problem of finding the muon trajectories ends up being finding the (local) maxima in them.

The HB algorithm does not use the Hough transform as-is, but a derivation from it called compact Hough transform (CHT). In it, the 2D problem is reduced to a 1D histogram. The details of this transformation and the implementation of the entire HB algorithm in an FPGA can be found in [144].

The preparations for the design of the L1 trigger of the Phase 2, that culminated in the release of the L1 TDR for CMS' Phase 2 [134], included projections of trigger efficiency (measured against reconstructed segments) with the ageing scenario described in [136]. To do so, both AM and HB algorithms were put to test with a MC sample simulated to mimic the collision environment of the HL-LHC, with 200 simultaneous collisions in average (i.e. $\mu = 200$). In addition, the same techniques to mimic the ageing mentioned in Sect. 2.4.1 (i.e. applying the reduction in hit efficiency as a probability for the detection of the signal), were used to apply the ageing scenario, also adding RPC failures. The main results can be seen in Fig. 3.3. The overall trigger efficiency comparison is calculated with respect to (locally) reconstructed segments that are geometrically (i.e. in the η-φ space) matched with particle-level muons that have at least 20 GeV of p_{T}. The comparisons are shown also, only for the AM algorithm, in [143].

The results show that both algorithms have similar efficiency without simulated ageing, with almost a 100%. The AM algorithm with the extra step of using RPC information, is also shown for the timing resolution and efficiency, with reasonably similar outcomes (slightly overall efficiency than "bare" AM and slightly better resolution). Both algorithms do not have comparable results for the resolutions of φ and φ_b, where AM clearly performs better.

The comparisons with the ageing scenario show a reduction in overall efficiency appreciable in the MB1 and MB4, that is more significant in the external wheels of the detector (± 2 essentially, and less for ± 1). The largest reductions appear for MB1 in the ± 2 wheels, with a decrease of $\approx 45 - 50\%$ for the HB algorithm, and $\approx 30\%$ for the AM. The fact that the reductions appear in these stations is understandable, as they are the closest to the beam (the MB1s). The wheels corresponding to a larger inclination (η) from the interaction point of CMS (± 2) are also those more affected by radiation from particle scatterings (leaving aside the endcaps, of course): thus, it is logical as shown in Fig. 2.17, that these DT chambers are more affected by ageing. MB4s also should be more affected by background radiation, such as neutrons, that might scatter in the CMS' cavern, leaving the MB2 and MB3 as those with less ageing effects.
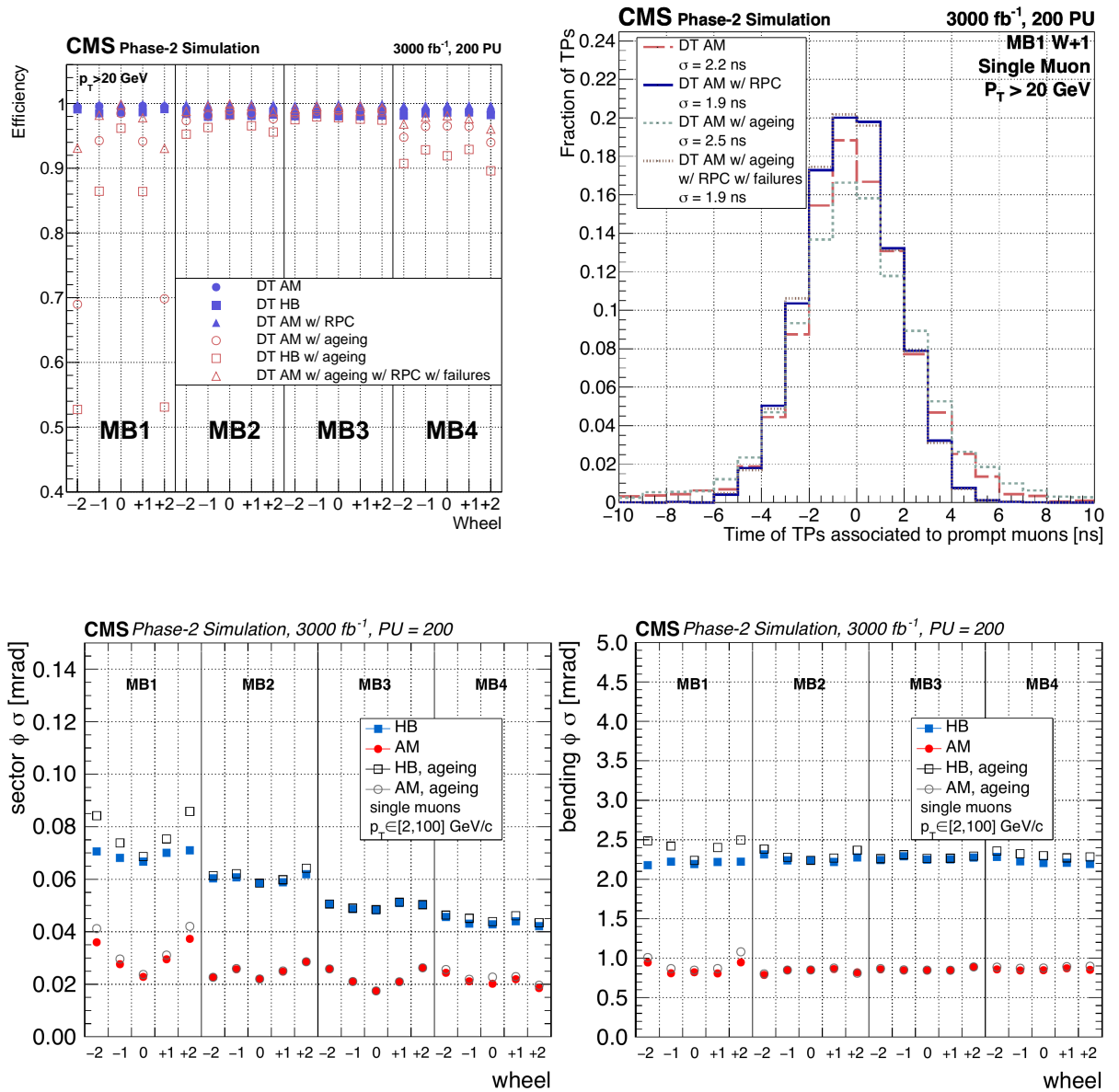
Figure 3.3: [134] Comparisons of the performance of TP-generating algorithms designed for the Phase 2 L1 trigger of CMS' DT subdetectors. Plots show the AM and HB algorithms, with values corresponding to their operation with and without the ageing scenario depicted in [151], in some cases simulating also failures in the RPCs. All observables are calculated with a MC simulation with $\mu = 200$ with muons with $p_T > 2$ or 20 GeV depending on the plot. The figures of merit shown are the trigger efficiency depending on the MB station and detector wheel (top left), the timing resolution in the MB1 of the +1 wheel (top right), and the φ resolution (bottom left) and the φ_b resolution (bottom right) depending on the station and wheel.

The figure also shows that a significant part of the efficiency lost by ageing of both DTs and RPCs can be recovered, in the case of the AM algorithm, by using RPC information. This reduces the efficiency looses to less than 9% in the worst cases of MB1s and less than 4% for MB4s. This also applies to the variation in the time resolution. Differences in φ and φ_b resolutions are analogously significant in MB1 and MB4 and especially for the ± 2 wheels. Ageing effects are expected to be larger for φ resolution.

3.1.3 Proposal for a Hough transform-based trigger primitive algorithm for the L1 muon trigger

In the context of preparations for the Phase 2 of CMS, I developed a proposal for an alternative algorithm to the mentioned AM and HB options to try to exploit even more the possibilities that the new Phase 2 electronics offer. This was done in parallel to another alternative algorithm, based slightly in a Bayes classifier called pseudo-Bayes algorithm, and detailed in [152].

The idea is to use the "pure" Hough transform, not the CHT, to extract the trajectories from the hit information and obtain the best candidates. Then, they are fed to the AM parameter-estimation component, to calculate φ and φ_b, taking advantage of its good spatial resolution. As mentioned before, using the Hough transform modifies the paradigm, being a local maxima search the challenge to be solved. Such algorithm can be extended a priori trivially to include hit information from RPCs. Although this can be also done by the default AM algorithm, it is important to note that it adds an extra step to do so, creating new TPs in the process. In the case of this proposal, it would be fully integrated in the same steps.

The steps of this new proposal are the following. Various steps (especially the last one) are in synchronisation with requirements also imposed on the pseudo-Bayes approach from [152]. For each MB DT chamber, and its digis:

1. **Travel to line space** The hits of both φ SLs, i.e. SL1 and SL3, are the inputs (as in the case of the other approaches to the issue) to the algorithm. The first step is to build the point/direct (x, y) representation of the hits collection of the chamber. To do so, each signal in one cell is duplicated in order to take into account the lateralities (left, right with respect of the cell's wire, see Fig. 2.10). Those (artificial) two hits are put equidistantly from the left and right cathodes, and the anode wire. Thus, they represent the middle point of each half (left and right) of one drift tube. For each digi, a 2D histogram describing the $\rho - \theta$ dual/line space is filled, by running $\rho = x \cos \theta + y \sin \theta$ between $\theta = -\pi/2$ and $\theta = \pi/2$. However, limits are imposed so that only a subset of θ is used, corresponding to a maximum physical inclination for any muon coming from the interaction vertex ($\approx 73^o$). The 2D histogram has a bin width in position identical to half the width of one cell (2.1 cm) and of one sexagesimal degree for θ bins. In Fig. 3.4 an example of the transformation of an DT chamber hit set (simulated) is shown.
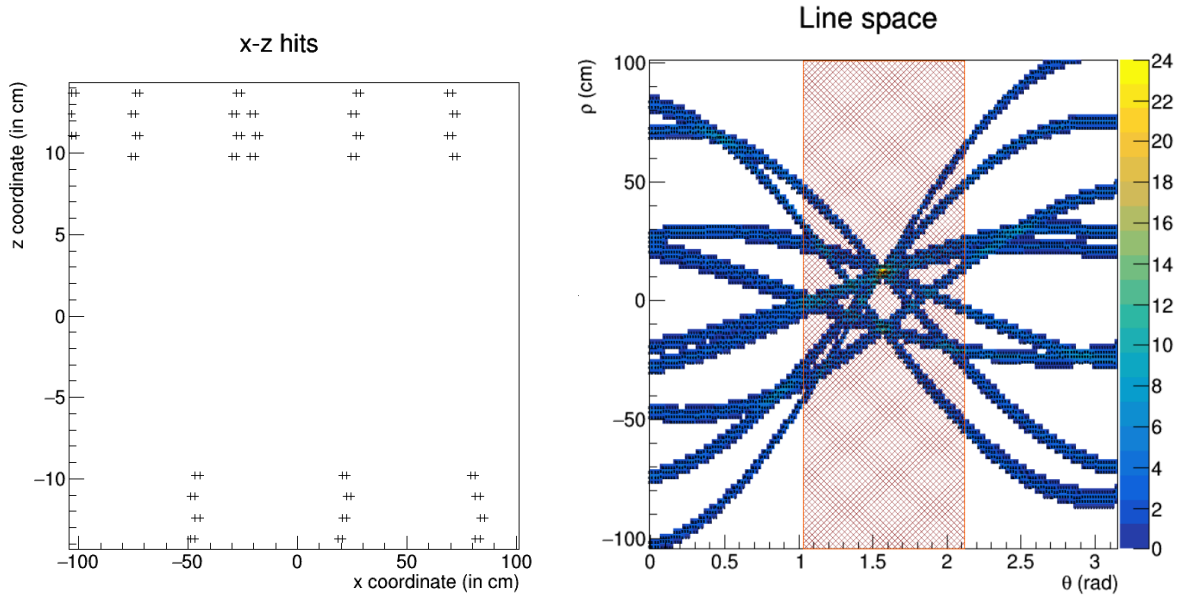
Figure 3.4: Representation of a simulated DT chamber set of hits in the direct/point space (left) and its equivalent in the dual/line space (right). Hits in the direct/point space have been duplicated to account for laterality information into the algorithm. The hatched red band shows the forbidden dual/line subspace, according to the maximum physical inclination a muon from the interaction vertex should have.

2. **Search of local maxima**

    a) The positions (bins) that contain more than a set threshold are extracted from the histogram. If no element passes it, it is lowered various times until a fixed minimum is reached.

    b) The entries coming from the previous step are clustered, by maximising $\Delta\rho$ and $\Delta\theta$. Each cluster is a set of bins.

    c) For each identified cluster, a weighted centre $(\rho_{\max}, \theta_{\max})$, that depends on the entries of each bin is derived. The values of a straight line in the direct/point space are recovered by inverting the Hough transform with $(\rho_{\max}, \theta_{\max})$. An example of such a line can be seen in Fig. 3.5.

3. **Hit association** For each line in the point space obtained from the dual one, we attempt to associate hits (and lateralities) to it and build a muon protocandidate. For each of the eight layers of SL1 and SL3, the existence of a hit in the cell that the line crosses is checked. If a hit exists, it is associated to the line, and lateralities are assigned depending on the location (to the left or right of the wire). If the line crosses very close to the wire of the cell (less than 0.03 cm), no laterality is assumed. In the case that no hit is present in the crossing cell, the left and right neighbours are checked, and lateralities are assigned then accordingly (right and left, respectively). One hit can be associated to more than one protocandidate.

4. **Ordering** Before feeding the final muon candidates to the AM parameter-extracting methods, the protocandidates are pruned to avoid spurious elements, duplications, or low quality entries. First of all, protocandidates are ordered depending on (hierarchised classification criteria, i.e. if two elements are not ordered by one of them, the next one is applied):

    a) Number of layers with hits.

    b) Number of hits in the exact cell that the line crosses.

    c) Number of hits in adjointing cells.

    d) Absolute difference of number of hits between SL1 and SL3.

    e) Sum of the absolute distances between the line that generated the candidate and the wire (if no laterality is set) or the middle of half a cell (if laterality is set).

   Then, if any protocandidate shares hits with another, that of less value (depending on the order just described) looses them. If the shared hit has an opposite laterality, the better candidate keeps the hit, but looses the laterality. The remaining set of segments without shared hits are called muon candidates.

5. **Filtering** After this step, quality criteria are imposed to the muon candidates before sending them to the AM algorithm. These requirements are e.g. having a minimum number of layers with hits, requiring any combination of N hits in the SL1 and N' in SL3 (i.e. 3+3, 3+2), allowing candidates with associated hits in only one SL, etc..

The algorithm has been implemented, as well as the other three proposals, in a `C++` emulator of the trigger, that will run on an FPGA. Despite its simplicity, it had some adversities to face. The main issue was the local maxima finding: a known mathematical challenge, but that needed to deal with the hardware constraints of the L1 trigger. These comprised mainly latency as well as the implementation in hardware. Currently, CMS efforts for the new DT L1 trigger are focused on the AM algorithm.

x-z hits
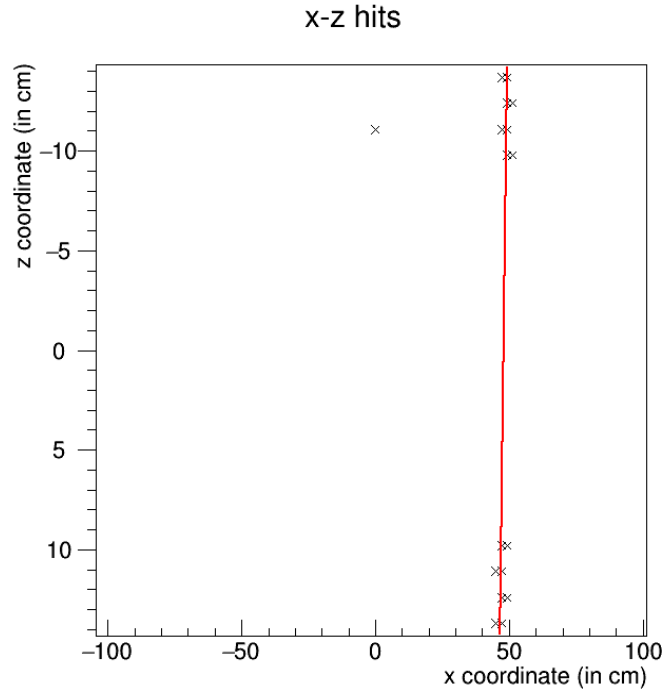
Figure 3.5: Representation of a simulated DT chamber set of hits in the direct/point space, duplicated to account for laterality information into the algorithm. One line, derived from the maxima search step of the proposed algorithm, is superimposed.

3.2 Generation

We commonly call "generation" the process of obtaining MC samples to confront predictions from the SM (or other theories) with data. One could separate the process into three different steps, following the overview from the beginning of the chapter and the details from Sect. 1.3.1: the generation of MC-based **matrix element** (ME) final state particles four-momenta, their "showering" in the **parton shower** (PS), and the simulation of how these particles would spread from the collision point of CMS and **interact** with its subdetectors.

The software that is able to do the first step (ME) is commonly called "generator", and currently the most used in the field are `Pythia 8` [29], `POWHEG` [153, 154, 155], `MadGraph5_aMC@NLO` [156] and `Sherpa` [157]. `Pythia 8` is able to generate process only at LO, `MadGraph5_aMC@NLO` is able to run processes at LO and reaches also NLO, as well as `POWHEG` and `Sherpa`.

As explained in Sect. 1.3.1, to do accurate simulations to data we must take into account the PS. To do so, we usually get the output of a ME software, and then use it as input to another program that can shower its final state particles. The most knowns are `Pythia 8` (also a ME generator) and `Herwig 7` [158, 159]. These PS programs have several parameters that allow to adjust how they perform over while showering/hadronising. Commonly, the parameters are fixed or **tuned** by fitting to specific data, such as e^+e^- collisions with multiple jets, or pp collisions from previous Runs with lower energy.

Finally, for the last step of simulating how the particles travel the different subdetectors and make energy depositions and signals in them, we use for most MC samples the `Geant4` [160, 161, 162] software. `Geant4` is a general particle-matter interaction program that is used in many different scopes apart from high energy physics, such as medical applications, or to take into account the interaction of cosmic ray radiation into spaceships.

3.2.1 Differential comparisons

Although the value of the cross section itself can be compared with predictions as we mentioned, BSM theories can also affect the event distributions themselves (or even, they could have a small or negligible effect in the normalisation, but a large one in the variables' shape). This is one of the reasons to study the modelling of distributions, by measuring the cross section differentially, i.e. depending on one (or various) observables (e.g. the energy of one final state particle). There are other motivations to do so, such as studying the modelling of different MC generators: improving them is crucial in order to do better measurements, as we mentioned before. Measuring differentially observables can be tricky, as the efficiency and acceptance that one should take into account are now not scalars, but matrices (or, in more than one dimension, tensors). The methodological aspects are detailed later in Sect. 7.1.

Once we overcome these difficulties, where we remove the detector noise and effects from our data, we arrive to what is dubbed "generator level", that is, we can compare directly with generated information, without the need of simulating how their final state particles travel through the detector and reconstruct afterwards them. However, as the reality of the proton-proton collision is a continuous sequence of interactions (among which are those of the process of interest or signal), there are two generator "levels" defined in the field, that must be differentiated. The first one, **parton level**, is defined by the information of the particles produced after the generation of the hard-scattering process, whereas the **particle level** corresponds to the particles after the parton shower and hadronisations.

There is a "common ground" to define the details of these levels that is the `Rivet` [163] toolkit: a complete computing framework to preserve previous (and future) differential analyses from various experiments and collaborations, allowing to run their "routines" (their analysis code) with different generators to compare with their data and the same definitions of objects the modelling of the distributions. For the studies described in this thesis, we will unfold results to particle level using `Rivet` definitions of particle-level objects. Due to the showering and fragmentations, care must be taken with quarks and leptons, as the first ones have already produced a jet of particles, and the second ones could suffer QED radiation. Consequently, `Rivet` definitions usually include some recommendations to reconstruct jets at particle level and also to re-define the particle level four-momentum of leptons by adding to it the four-momenta of collinear photons.

The definitions commonly used in the CMS Collaboration at particle level are the following, also used in the studies of this document. Objects are constructed using stable (i.e. with a

lifetime larger than 30) generated particles following the conventions given in Ref. [164], where these definitions are explained in detail. Muons and electrons not coming from hadronic decays (prompt leptons) are "dressed" by taking into account the momenta of nearby photons within a $\Delta R < 0.1$ cone, where ΔR is the separation in η-φ space between the muon or electron and the photon. Jets are clustered from all of the stable particles excluding prompt electrons, prompt muons, prompt photons, and neutrinos, using the anti-k_t algorithm with a distance parameter of $R = 0.4$. The information of the intermediate hadrons and τ leptons is preserved inside the jets and used to determine whether a jet originates from the fragmentation of a heavy-flavour quark (bottom or charm) or whether it is a decay product of a τ lepton.

3.3 Simulations used

For simulations, several are used to model the signal and background processes. Background ones are estimated using only simulation samples, all of which are listed in Tab. 3.1. As there are multiple, very different, processes, we group them in four sets: first the top quark pair-production decaying dileptonically, $t\bar{t}$. Then, the Drell-Yan process or DY. Events that might have two leptons correctly reconstructed and identified, but in which one (or more rarely, both) comes actually from a jet (a *fake* lepton) are collected in the Non-W/Z group (as the leptons do not arise from a W or Z boson decay). Finally, the rest of the processes except the signal, mostly from dibosonic production, and $t\bar{t}$ associated production with a vector boson, are put in the VV+$t\bar{t}$V. The generated samples for tW and $t\bar{t}$ and the most of the rest of background processes use the NNPDF3.1 PDF set [165]. All tW and $t\bar{t}$ samples use the CP5 tune [166] to model the UE, as well as all other backgrounds in the 2017 and 2018 simulations. In 2016, the remaining background processes use the CUETP8M1 tune [167].

3.3.1 Alternative samples for uncertainty estimation

In addition to all the samples listed in the mentioned tables, others were used to estimate some uncertainty sources by varying parameters in their generation. These can be, for example, the mass of the top quark: these simulations are the same as the nominal ones except those parameters varied, as detailed later in Sect. 5.

Another example are those used to estimate the uncertainty linked to the colour reconnection models used by the PS modelled by `Pythia 8`. These are extensively detailed in [168]. Nominal samples use the standard approach, called "MPI-based" colour reconnection model, in which groups of MPI particles are assigned a probability to be colour-connected with others depending on its hardness scale (usually p_T) and other adjustment parameters. This is done under the assumption that MPI groups with large energy (p_T) would tend to flee the interaction point, and thus they will not be likely colour-connected with the hard-scattering particles. Then, colour strings (i.e. colour charge assignments to particles according to the shower) are defined, initially

Process (decay)	Generators	MC precision	σ (pb)	Norm. precision	Group
$t\bar{t}$ (dileptonic)	POWHEG +Pythia 8	NLO+PS	88.3	NNLO+NNLL	$t\bar{t}$
Drell-Yan (dileptonic, $m_{l\bar{l}} \in [5, 50)\,\mathrm{GeV}$)	MadGraph5_aMC@NLO +Pythia 8	NLO+PS	22635.1	NNLO+NNLL	DY
Drell-Yan (dileptonic, $m_{l\bar{l}} \geq 50\,\mathrm{GeV}$)	MadGraph5_aMC@NLO +Pythia 8	NLO+PS	6025.2	NNLO+NNLL	DY
$t\bar{t}$ (semileptonic)	POWHEG +Pythia 8	NLO+PS	365.4	NNLO+NNLL	Non-W/Z
WW (dileptonic)	POWHEG +Pythia 8	NLO+PS	12.2	NLO+PS	VV+$t\bar{t}$V
WW (semileptonic)	POWHEG +Pythia 8	NLO+PS	45.53	NLO+PS	VV+$t\bar{t}$V
WW (double scattering)	Pythia 8	LO+PS	1.6	LO+PS	VV+$t\bar{t}$V
WZ (trileptonic)	POWHEG +Pythia 8	NLO+PS	4.4	NLO+PS	VV+$t\bar{t}$V
WZ (dileptonic)	MadGraph5_aMC@NLO +Pythia 8	NLO+PS	5.6	NLO+PS	VV+$t\bar{t}$V
WZ (monoleptonic)	POWHEG +Pythia 8	NLO+PS	10.7	NLO+PS	VV+$t\bar{t}$V
ZZ ($\ell\ell\nu\nu$)	POWHEG +Pythia 8	NLO+PS	0.6	NLO+PS	VV+$t\bar{t}$V
ZZ ($\ell\ell qq$)	MadGraph5_aMC@NLO +Pythia 8	NLO+PS	3.3	NLO+PS	VV+$t\bar{t}$V
$t\bar{t}$W ($W \to \ell\nu$)	MadGraph5_aMC@NLO +Pythia 8	NLO+PS	0.2	NLO+PS	VV+$t\bar{t}$V
$t\bar{t}$W ($W \to qq$)	MadGraph5_aMC@NLO +Pythia 8	NLO+PS	0.4	NLO+PS	VV+$t\bar{t}$V
$t\bar{t}$Z ($Z \to \ell^+\ell^- \cup Z \to \nu\bar{\nu}$, $m_{\ell\ell/\nu\nu} \in [1, 10)\,\mathrm{GeV}$)	MadGraph5_aMC@NLO +Pythia 8	LO+PS	0.05	LO+PS	VV+$t\bar{t}$V
$t\bar{t}$Z ($Z \to \ell^+\ell^- \cup Z \to \nu\bar{\nu}$, $m_{\ell\ell/\nu\nu} \geq 10\,\mathrm{GeV}$)	MadGraph5_aMC@NLO +Pythia 8	NLO+PS	0.3	NLO+PS	VV+$t\bar{t}$V
$t\bar{t}$Z ($Z \to qq$, $t\bar{t}$ dileptonic)	MadGraph5_aMC@NLO +Pythia 8	NLO+PS	0.1	NLO+PS	VV+$t\bar{t}$V
$t\bar{t}\gamma$ ($t\bar{t}$ dileptonic)	MadGraph5_aMC@NLO +Pythia 8	NLO+PS	1.5	NLO+PS	VV+$t\bar{t}$V
$t\bar{t}\gamma$ ($t\bar{t}$ semileptonic)	MadGraph5_aMC@NLO +Pythia 8	NLO+PS	5.0	NLO+PS	VV+$t\bar{t}$V
WWW (inclusive)	MadGraph5_aMC@NLO +Pythia 8	NLO+PS	0.2	NLO+PS	VV+$t\bar{t}$V
WZγ (inclusive)	MadGraph5_aMC@NLO +Pythia 8	NLO+PS	0.04	NLO+PS	VV+$t\bar{t}$V
WWγ (inclusive)	MadGraph5_aMC@NLO +Pythia 8	NLO+PS	0.2	NLO+PS	VV+$t\bar{t}$V
WWZ (inclusive)	MadGraph5_aMC@NLO +Pythia 8	NLO+PS	0.2	NLO+PS	VV+$t\bar{t}$V
WZZ (inclusive)	MadGraph5_aMC@NLO +Pythia 8	NLO+PS	0.06	NLO+PS	VV+$t\bar{t}$V
ZZZ (inclusive)	MadGraph5_aMC@NLO +Pythia 8	NLO+PS	0.01	NLO+PS	VV+$t\bar{t}$V

Table 3.1: MC simulations used in the inclusive and differential measurement of the tW process to model background processes. Each sample shows the associated process, the generators used, the precision of the simulations, the normalisation cross sections used, the precision of this value, and the group in which they are put in the analysis.

depending on the p_T, and finally by a minimisation of the total colour strings length.

One alternative model used only diverges from the nominal by allowing the top quark or W boson decays to participate in the CR procedures. Usually, the algorithm reconnects colour between MPI groups before decaying these particles: the early resonant decays (ERD)-activated model "ERDon" allows them to participate. Another forces the rules of QCD when defining the colour strings, and thus is called "QCD-based". It thus causally connects all particles in the strings in spacetime. The third approach considered in this analysis, called "gluon-move based model", allows that final-state gluons from the hard process to "move" across strings. This can help reducing the total string length.

4 Object identification and corrections

W ITH the products of the reconstruction from the previous section, we have the ingredients to do our analyses with them. However, it is usually common to enhance the identification requirements done in the reconstruction stage and demand more qualified objects. As this is standard in most of the analyses, the CMS Collaboration develops common identification criteria for leptons, photons and jets. In this thesis we will use these central provided object identification criteria.

4.1 Electrons

For electrons we use the dedicated central identification (ID), with the most strict or "tight" recommended settings (or working point, WP). The recommended ID criteria are built upon PF electrons, reconstructed as detailed in the previous section. This selection allows for an efficiency of identification of $\approx 70\%$. There is a dependence on the η of the associated supercluster (η_{SC}), as particles that have large η travel through different subdetectors (partly those of the endcap) than those more centrally produced. This, for the case of electrons, can be clearly understood by looking the disposition of the tracker and ECAL as the quadrant shown in Fig. 4.1. The two ECAL components, the barrel's (EB) and endcaps' (EE) are separated at an inclination of $\eta = 1.479$ from the centre of the detector. The requirements from this ID applied to this document's analyses electrons are shown in Tab. 4.1. Apart from them, two extra criteria are used: first, a minimum p_{T} of 20 GeV. Secondly, that they must have $|\eta| < 2.4$ and they must not lie in the transition region between the EB and EE, i.e. $1.444 < |\eta| < 1.566$. This is done to avoid this area where the reconstruction is not so efficient, as shown in Fig. 2.12. Also, an additional condition on the impact parameter with respect to the PV is imposed, requiring less than 0.05 (0.10) mm in d_{xy} for electrons reconstructed in the barrel (endcap) ECAL and less than 0.10 (0.20) mm in d_z accordingly.

Details on the optimisation and the selection of variables can be found in [119]. However, a brief description follows. $\sigma_{i\eta i\eta}$ is the second moment (i.e. standard deviation) of the distribution (log-weighted) of the collected energies in the ECAL depending on η, calculated around the most energetic ECAL crystal cell from the SC and scaled to crystal size units. $\Delta\varphi_{in}$ is defined as the difference in φ among the SC and the track. In a similar way, $\Delta(\eta_{\mathrm{in}}^{\mathrm{seed}})$ is defined as the difference between the cluster seed rapidity and the track one (extrapolated from the innermost
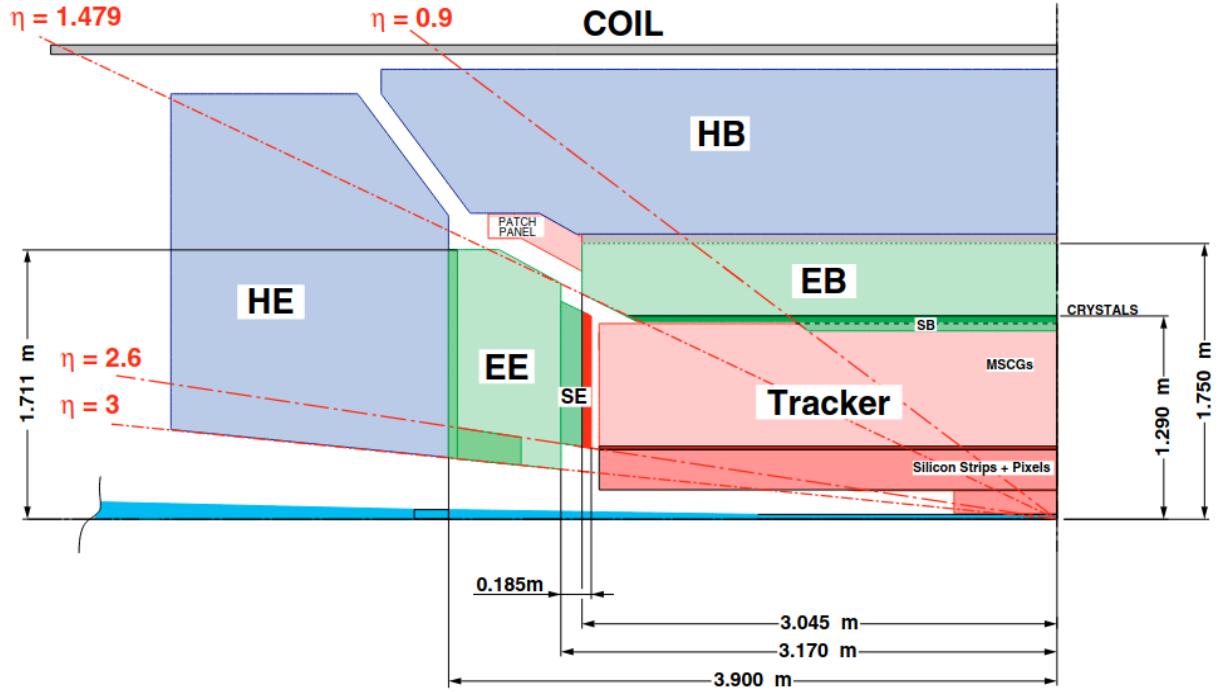
Figure 4.1: Diagram of one CMS quadrant near the beam axis. Tracker is shown in red near the LHC pipe. ECAL, separated in its barrel (EB) and endcap (EE) parts, appears in green tones. HCAL, analogously as ECAL in its HB and HE components, is displayed in blue.

Variable	$|\eta_{SC}| \leq 1.479$	$|\eta_{SC}| > 1.479$
$\sigma_{i\eta i\eta}$	< 0.0104	< 0.0353
$|\Delta(\eta_{in}^{seed})|$	< 0.00255	< 0.00501
$|\Delta\varphi_{in}|$	< 0.022	< 0.0236
H/E	$< 0.05 + 1.16 \cdot \frac{1}{E_{SC}} + 0.0324 \cdot \frac{\rho}{E_{SC}}$	$< 0.0155 + 2.54 \cdot \frac{1}{E_{SC}} + 0.183 \cdot \frac{\rho}{E_{SC}}$
I_{rel}^{EA}	$< 0.026 + \frac{0.0506}{p_T}$	$< 0.0453 + \frac{0.0963}{p_T}$
$|E^{-1} - p^{-1}|$	< 0.159	< 0.0197
Missing inner hits	≤ 1	≤ 1
Pass conversion veto	Yes	Yes

Table 4.1: Main electron identification requirements. They correspond to the tight WP centrally recommended by the CMS Collaboration.

track position). H/E is the ratio between the energy deposited in the HCAL and that of the electron candidate, in a cone of $\Delta R = 0.15$ around the SC direction. This requirement is made dependant on the energy of the supercluster (E_{SC}) and the pileup factor ρ as the energy deposited in the HCAL significantly changes with these variables. ρ is the median of the transverse energy density per area unit of the event. $I_{\mathrm{rel}}^{\mathrm{EA}}$ is the isolation defined in a cone of $\Delta R = 0.3$ by summing the transverse momenta from charged and neutral hadrons and photons upon which a correction is subtracted, and dividing the result by the particle's momentum (hence its "relative" alias). This correction, ρA_{eff}, depends on the pileup directly through A_{eff}: the effective area, defined as the area of the isolation region weighted so that it takes into account the pileup transverse energy density depending on the particle's η. The "missing inner hits" refer to the number of tracker crystal signals that can be expected to be found, according to the track, but they are not. This is a powerful discrimination variable against **photon conversions**: a process in which an original photon interacts (e.g. with the tracker's material) producing an electron and a positron. In these situations, our reconstruction might recover the information from these electron and positron, that are produced *after* the photon travelled a bit through the detector, thus not leaving signals in the tracker before interacting. However, as this ID criteria are dedicated to identify particles that directly come from the collision ("prompt" particles, as mentioned before), demanding a low number of missing hits in the innermost layers of the tracker allows to reduce contamination from photon conversions. The interaction point of photon conversions is reconstructed by PF. The "conversion veto" criteria is orientated actually towards photon identification: in it, it is demanded that no pixel detector seed of at least two hits in the pixel detectors can be matched to an ECAL cluster from a potential photon. This veto can be applied, inversely, for electrons: demanding that there is such a track linked to the electron's ECAL SC. Identification efficiencies of central electrons for a looser WP are shown in Fig. 4.2, as a function of their p_{T}. They have been estimated with the tag-and-probe method.

The reconstruction SF for electrons used in the analysis are shown in Fig. 4.3, whereas the corresponding for the ID are in Fig. 4.4.
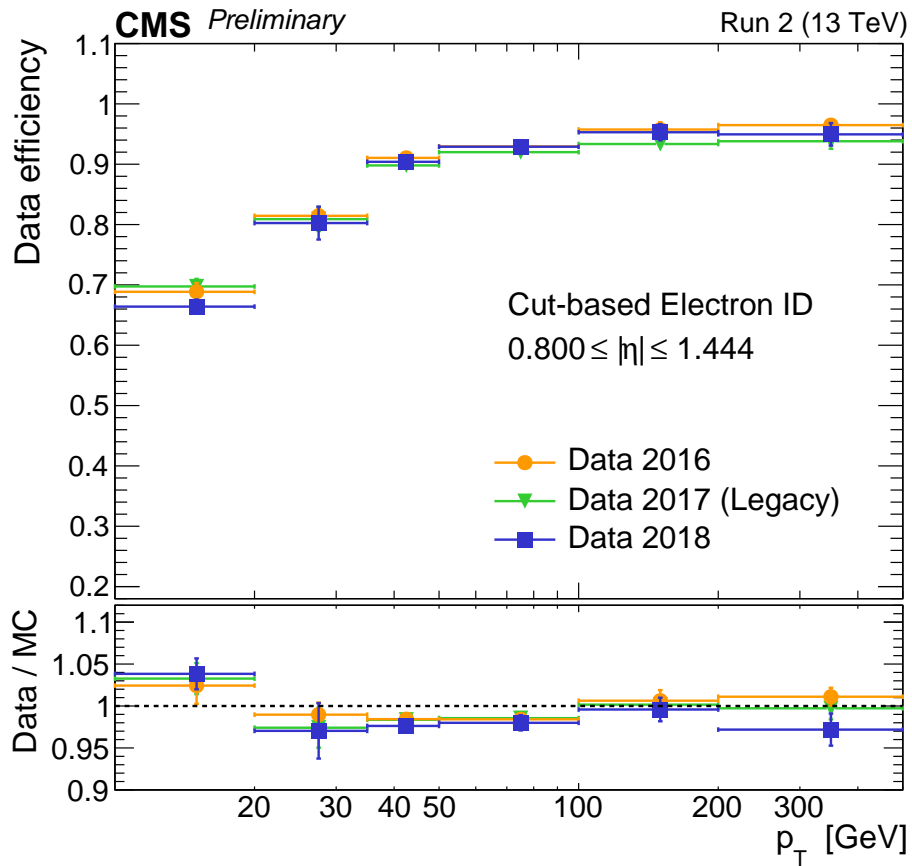
Figure 4.2: [169] Efficiency of electron identification for 2016, 2017, and 2018 data collected in proton-proton collisions with CMS at $\sqrt{s} = 13$ TeV for central electrons as a function of their reconstructed p_T. The ID requested uses the same variables as those of Tab. 4.1, but with a looser WP that achieves an overall 90% signal efficiency.

(a) 2016. (b) 2017. (c) 2018.
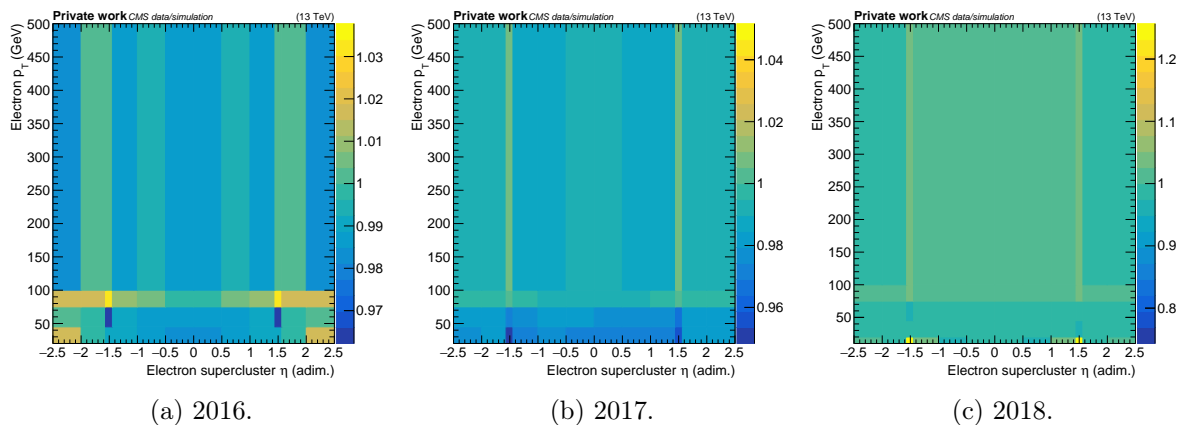
Figure 4.3: Electron reconstruction SF used for the different data-taking years.
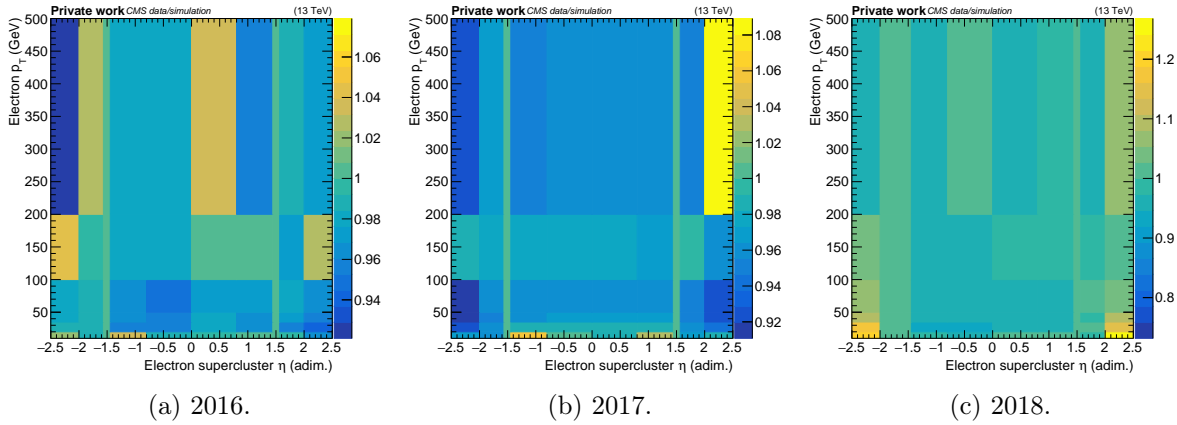
Figure 4.4: Electron ID SF used for the different data-taking years.

4.2 Muons

For muons, analogously as for electrons, we use the recommended central ID criteria with the tight WP, that allows for overall identification efficiencies larger than 96% for muons with $p_\mathrm{T} > 20$ GeV [170]. The criteria, imposed over PF-reconstructed global muons, are listed on Tab. 4.2. In the analysis, an additional requirement of $p_\mathrm{T} > 20$ GeV and $|\eta| < 2.4$ is imposed, as well as an isolation condition of $I_{rel} < 0.15$ in order to improve the identification of muons from W or Z boson decays. Detailed studies of the performance of this ID can be found in [170, 118]. A brief description of the variables used follows now.

χ^2 refers to the estimator value of the fit's minimisation to the global muon track, whereas N_dof to the number of degrees of freedom: imposing a cut to its ratio is imposing a cut on the detector signals' adequateness to a theoretical muon track. $N_\mathrm{hit}^\mathrm{DT}$ is the number of hits from muon chambers included in the global-muon track fit. $N_\mathrm{segment}^\mathrm{DT}$ is the number of chambers that have a matched segment with a tracker trajectory: this requirement forces that the muon is a tracker muon also. d_i^tracker is the impact parameter from the associated tracker's track projected in the i dimension. $N_\mathrm{hit}^\mathrm{pixel}$ is the number of hits in the crystal tracker pixels and $N_\mathrm{layers\ with\ hits}^\mathrm{tracker}$ the number

Variable	Criterium
χ^2/N_dof	< 10
$N_\mathrm{hit}^\mathrm{DT}$	≥ 1
$N_\mathrm{DT}^\mathrm{segment}$	≥ 2
d_{xy}^tracker	< 2 mm
d_z^tracker	< 5 mm
$N_\mathrm{hit}^\mathrm{pixel}$	≥ 1
$N_\mathrm{layers\ with\ hits}^\mathrm{tracker}$	≥ 6

Table 4.2: Main muon identification requirements. They correspond to the tight WP centrally recommended by the CMS Collaboration.

of tracker layers with hits. The efficiency of (tight ID) identification and reconstruction of muons with collisions at $\sqrt{s} = 13$ TeV are shown in Fig. 4.5, estimated with the tag-and-probe method.
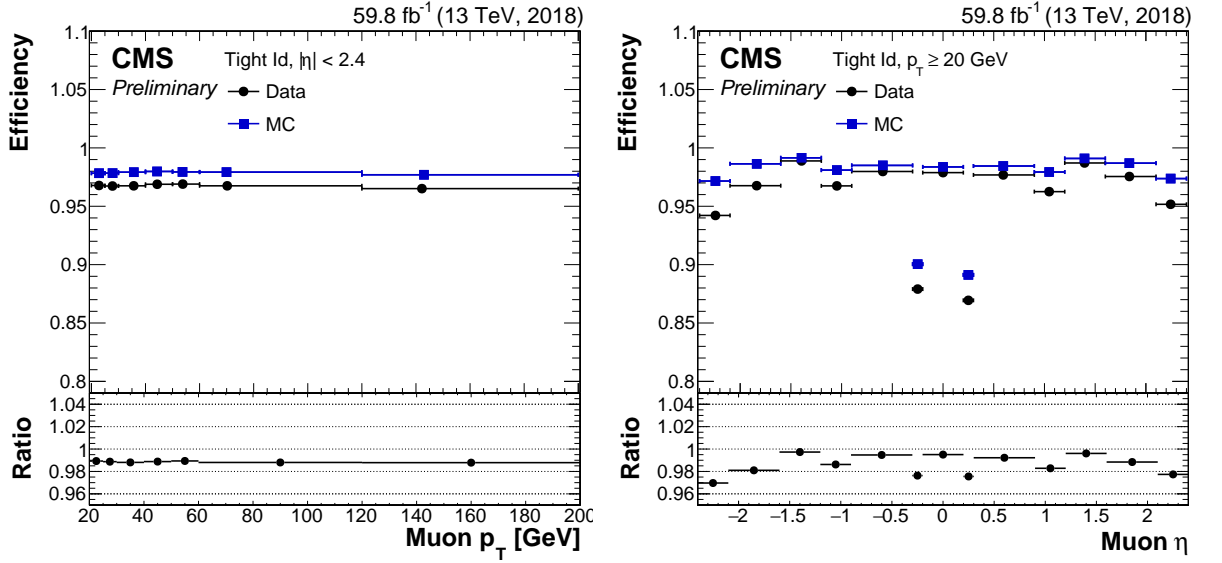
Figure 4.5: [171] Efficiency of reconstruction and identification of muons according with the criteria of Tab. 4.2 as a function of their p_T (left), and of their pseudorapidity (right). Data is collected from proton-proton collisions at $\sqrt{s} = 13$ TeV with muons of $|\eta| < 2.4$ for the p_T dependence and with $p_T > 20$ GeV for the pseudorapidity variation.

Muon reconstruction SF are very close to the unity and thus they are not applied. Muon identification and isolation SF are shown in Figs. 4.6 and 4.7 respectively: they are derived in the 2016 year in two periods of time, consisting of different data-taking blocks, because of a difference in the tracker detector (that affected efficiency).
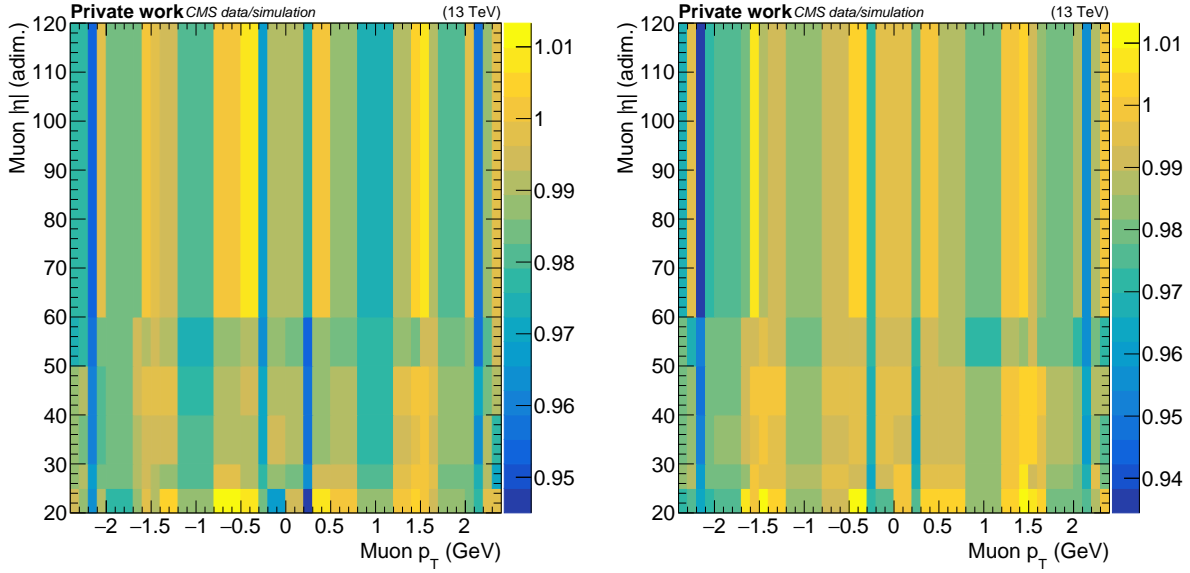
(a) 2016 first period.

(b) 2016 second period.
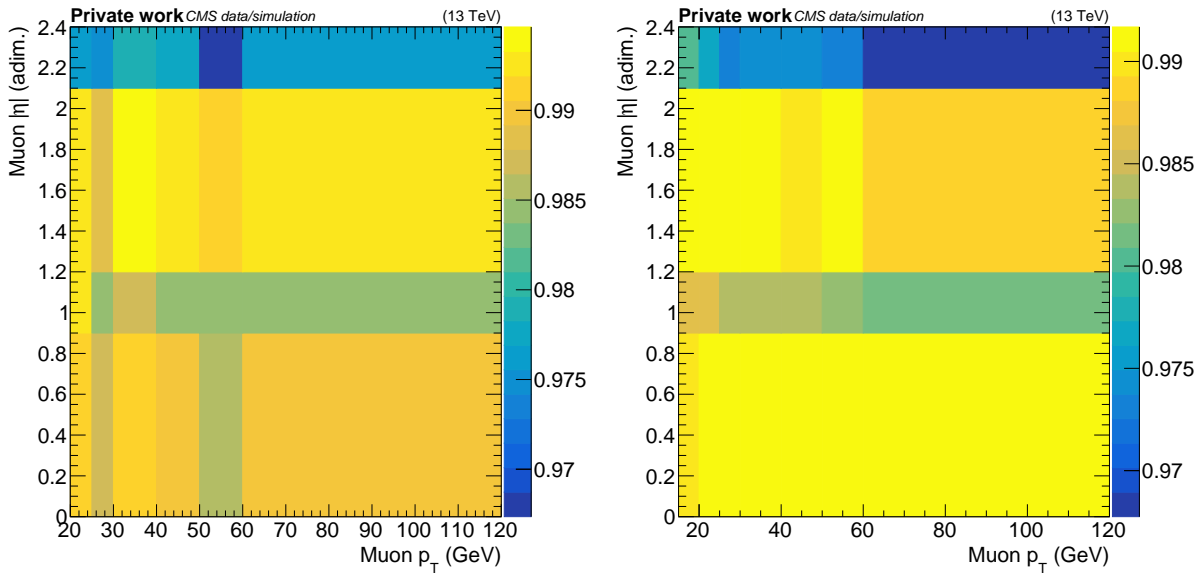
(c) 2017.

(d) 2018.

Figure 4.6: Muon ID SF used for the different data-taking years.

(a) 2016 first period.

(b) 2016 second period.

(c) 2017.

(d) 2018.
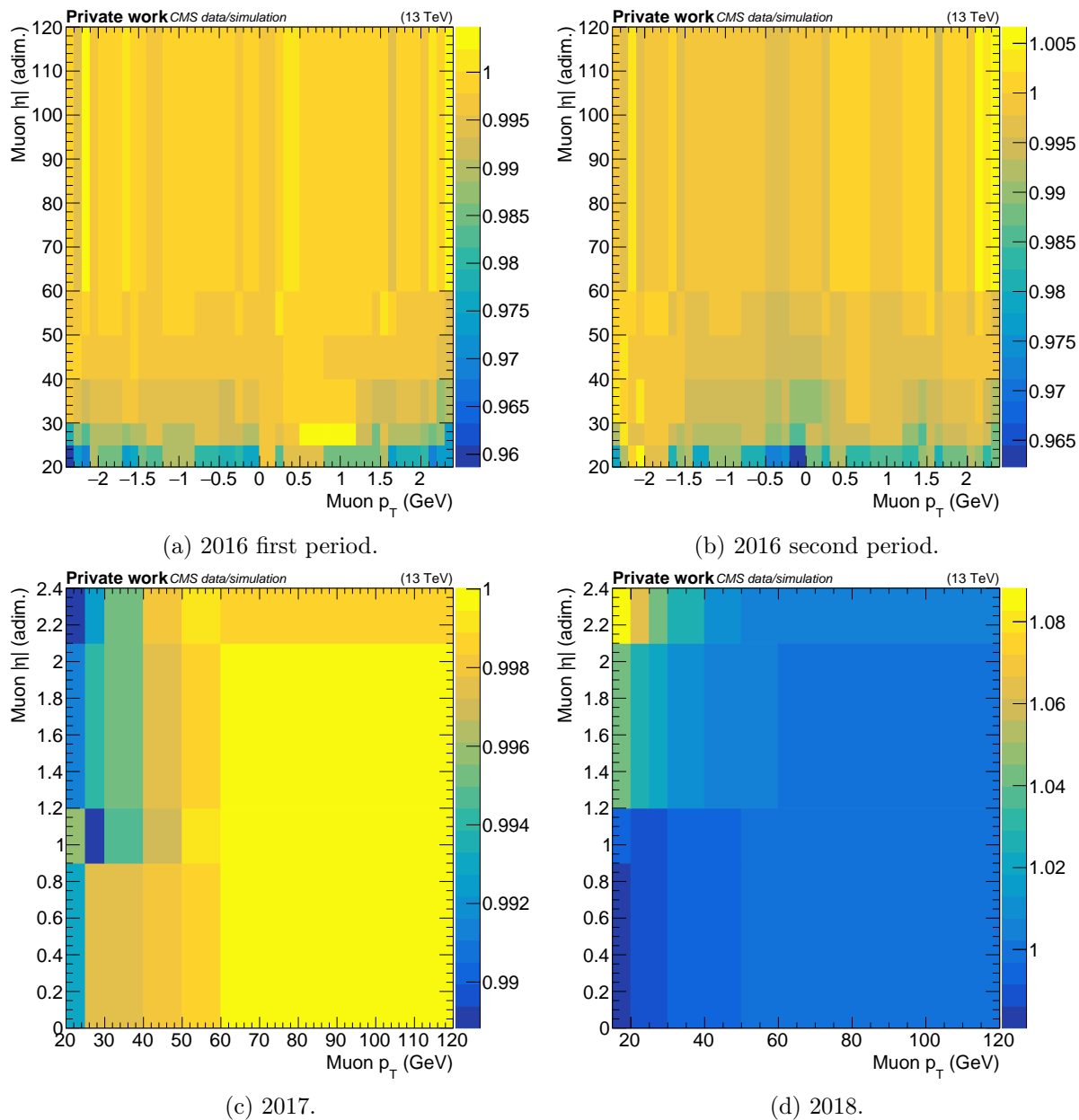
Figure 4.7: Muon isolation SF used for the different data-taking years.

4.3 Jets

Again, as for electrons and muons, we use the recommended central ID criteria with the tight WP, that allows for overall identification efficiencies $\approx 99\%$ when applied to PF-reconstructed jets. The criteria for the central ($|\eta| \leq 2.4$) jets used in the analysis are listed on Tab. 4.3. Detailed studies of the performance of this ID can be found in [172]. A brief description of the variables used in the ID follows now.

Variable	Criterium
NHF	< 0.9
NEMF	< 0.9
CHF	> 0
N_{charged}	≥ 1
$N_{\text{const.}}$	≥ 2

Table 4.3: Main jet identification requirements. They correspond to the tight WP centrally recommended by the CMS Collaboration with $|\eta_j| < 2.4$.

NHF is the neutral hadron fraction: the ratio of the energy of neutral hadrons inside the jet over the total energy of the jet. This definition applies analagously for CHF (charged hadron fraction). NEMF refers to the equivalent fraction but of the photons inside the jet. $N_{\text{const.}}$ is the number of elements that conform the jet, while N_{charged} is the number of those that are charged.

Jets in the analysis must have at least 30 GeV of p_{T} and $|\eta| < 2.4$. To ensure that the collection of jets is clearly not mixed with that of leptons, because of possible misidentification of either one jet or lepton, any jet that lies within a cone in the $\eta - \varphi$ space of $\Delta R < 0.4$ is removed. If one jet fulfills this ID but instead of $p_{\text{T}} > 30$ GeV has 20 GeV $< p_{\text{T}} < 30$ GeV, we put it into the **loose jets** category, while we call simply "jets" the rest of them. A b-tagging algorithm is applied upon jets, as described in 2.3.7.1, using a WP with ID efficiencies of $\approx 70\%$ and misidentification of $\approx 1\%$.

4.4 Corrections to simulations and data

If we were to compare our simulated events to data after their selection and filtering based on the previous reconstruction and identification methods, we might run into wrong conclusions. The effect that e.g. our muon identification criteria has might be different in data and in simulations, and a priori we do not know it. This is a direct consequence of our simulations not being perfect, corroborated by experimental results that indeed show difference in efficiencies.

To avoid this issue, a common procedure is to modify our samples by applying to the MC events weights a factor that accounts for the difference in efficiency (for the previous example, depending on the presence or not of identified muons with that ID) with data. This number is what we call a **scale factor** (SF).

However, there are other modifications that are applied mostly to the MC simulations information, and some even to data. In this section we review the modifications applied in this thesis' analysis.

4.4.1 Pileup

The number of average interactions per crossing (collision) varies not only depending on the accelerator conditions (as shown in Fig. 2.6, where it changes per year), but also between data and simulations. A reweighting is applied to the MC simulations, derived from the ratio of the average number of interactions (or pileup profile) in data and MC. From simulations, the distribution is easily obtained, as the amount of simultaneous interactions is known and derived for each bunch collision from a fully determined Poisson distribution. For data, things are different: obtaining the pileup profile is closely related to the calculus of the instantaneous luminosity. Here follows a quick summary, though a detailed explanation can be found in [103]. From Eq. (2.2),

$$N = L \cdot \sigma \;\Rightarrow\; N_{\text{pp interactions}} = L \cdot \sigma_{pp}^{in} \;\Rightarrow\; \mu\nu = \mathcal{L} \cdot \sigma_{pp}^{in} \;\Rightarrow\; \mathcal{L} = \nu\frac{\mu}{\sigma_{pp}^{in}}, \tag{4.1}$$

where μ is the average number of simultaneous collisions per bunch crossing (i.e. the pileup *plus* the interaction of interest[1]), ν the LHC collision frequency, and σ_{pp}^{in} the inelastic proton-proton collision cross section. We have returned in the lasts steps to the instantaneous luminosity ($\mathcal{L}$), so that we can speak in per-collision terms. Following the last identity, as ν is known, measuring the $\frac{\mu}{\sigma_{pp}^{in}}$ ratio in any way allows us to determine the instantaneous luminosity $\mathcal{L}$. There are various methods to estimate this quotient. One of the used in CMS is that of hit counting, used by the forward hadronic calorimeter (HF). To measure it, we transform Eq. (4.1) as

$$\mathcal{L} = \nu\frac{\mu}{\sigma_{pp}^{in}} \;\Rightarrow\; \mathcal{L} = \nu\frac{\mu_{\text{vis.}}}{\sigma_{pp,\text{vis.}}^{in}}, \tag{4.2}$$

where we use the fact that the ratio $\frac{\mu}{\sigma_{pp}^{in}}$ must be the same as the same quotient but in a restricted (*vis*ible) part of the entire phase space of the interactions (the one of the HF subdetector), i.e. the average pileup (plus collision of interest) in that space $\mu_{\text{vis.}}$ over the fiducial cross section there $\sigma_{pp,\text{vis.}}^{in}$: $\frac{\mu_{\text{vis.}}}{\sigma_{pp,\text{vis.}}^{in}}$. $\mu_{\text{vis.}}$ can be calculated as described in [103], and it is done in average in a period called a **lumisection**, that lasts ≈ 23 s and equals 64 luminibbles: another time period that corresponds to 2^{12} complete trips around the LHC circumference (or orbits) by a circulating proton [173]. The measurement of $\sigma_{pp,\text{vis.}}^{in}$ corresponds to what is commonly called the calibrations of luminosity, and depends on the detector and accelerator conditions, as well as the colliding particles (e.g. if they are protons, or protons and electrons). It can be measured through e.g. the known van der Meer scans method [174].

---

[1]For luminosity measurements, "pileup" is used to name the *all* simultaneous interactions per bunch crossing, whereas for physics analyses and in general in the HEP community it usually refers to the *other* collisions, apart from the the collision of interest. We will use as pileup the second meaning in this document.

The average visible number of collisions per bunch crossing $\mu_{\text{vis.}}$ distribution cannot be used directly to compare with that of our simulations. It corresponds to a restricted phase space, and the average number will be surely lower than the real (total) one. Thus, the inclusive average number of collisions per bunch crossing μ is derived from the instantaneous luminosity $\mathcal{L}$, using simply

$$\mu = \frac{\mathcal{L} \cdot \sigma_{pp}^{in}}{\nu}. \tag{4.3}$$

ν is known, as well as $\mathcal{L}$ (per lumisection) for all data events. The only element that we need is σ_{pp}^{in}, which is measured in dedicated analyses and found to be (68.6 ± 1.7)mb [175] at $\sqrt{s} = 13$ TeV as observed by the CMS Collaboration.

4.4.2 Trigger

Given that trigger requirements are a selection, forcely applied in data and simulated for MC, a correction is needed for the difference in efficiency as mentioned before: a scale factor or SF. The efficiencies needed to derive the SF in this document are calculated using a different set of trigger requirements than those of the analyses. With them a baseline, a common denominator, can be established in both data and MC. By using the SF derived through this technique, we are assuming that the difference in efficiency with respect to this reference trigger is equivalent to that of the (unmeasurable) inclusive efficiency between data and MC: to make this premise more bearable, the reference trigger is usually chosen to be as much uncorrelated with the HLT paths used, and as less demanding as possible. Then, efficiencies in both data and MC can be extracted for a particular phase space region R as

$$\varepsilon = \frac{N\,(\text{Passes the reference trigger} \wedge \text{Passes the analysis trigger} \wedge \text{Is inside } R)}{N\,(\text{Passes the reference trigger} \wedge \text{Is inside } R)}, \tag{4.4}$$

and with the ratio between data and MC, the SF can be derived. As the trigger depends on information from various subdetectors quickly processed, the performance of these systems might vary when checked against one or other variable. The SF used in the analysis depend on the p_{T} of the two leading leptons identified, that are expected to be an electron and muon. The values are shown in Fig. 4.8.

4.4.3 Lepton

In both reconstruction and identification a selection is done, therefore the efficiencies in data and MC can be different, and SF might be needed. But, in addition, the measurement of the energy of the particles is usually corrected from the raw values obtained from reconstruction to account for known deviations from the subdetectors. These corrections in energy can be applied to both

(a) 2016. (b) 2017. (c) 2018.
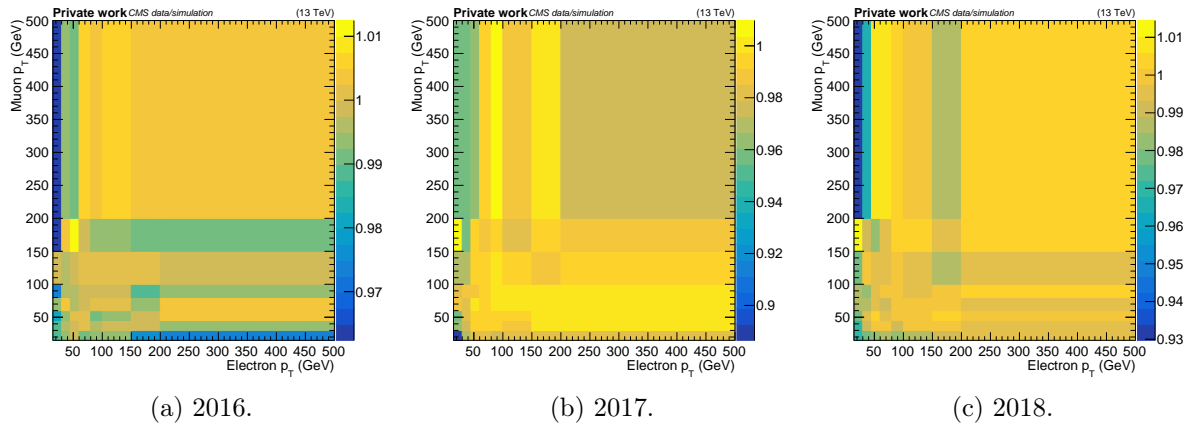
Figure 4.8: Trigger SF used for the different data-taking years.

data and MC.

4.4.3.1 Reconstruction and identification

Efficiencies in the reconstruction and ID of leptons can be estimated through the mentioned tag-and-probe method, and be put as a function of the particles' properties, such as their η or p_T (most commonly). This allows to extract SF between data and MC that can be used later to correct the simulations to data.

In the case of muon reconstruction, the differences in efficiencies between data and MC are so small than the SF derived are usually almost one in all cases. Thus, their are commonly not used. This does not apply to electrons: due to their particle-matter interaction features in the energy regime of high energy collisions at the LHC, the differences are slightly larger.

In the case of ID criteria, for muons there is a separation of efficiencies in the requirement for isolation: usually, the efficiency of the criteria and the efficiency on the isolation requirement are estimated separately to obtain SFs. This is done because of the potential difference of the jet properties and components in different areas of the entire phase space. Thus, two SFs are derived for the ID of muons, whereas for electrons only one SF is estimated.

4.4.3.2 Energy

The energy of muons is corrected to take into account bias from the subdetectors using the Rochester method [176], in which the well-known Z boson mass is used as anchor through its $\mu^-\mu^+$ decay (i.e. the Drell-Yan process). This provides corrections for all reconstructed muons in both data and MC.

For electrons, two corrections are applied to both data and MC, that are thoroughly detailed in [119]. First, the same simulation-tuned algorithm is run for both data and MC to exploit the maximum information possible from the detector so that the energy resolution is the best

achievable. This "energy regression", as it is called, does not remove the separation in energy scale or resolution between data and MC: it only improves the energy precision.

The second correction tries to amend this gap, by smearing the energy resolution in MC to match that of data, and by modifying the energy scale in data to that of MC. This is done using as reference process (or "candle") the Z boson decay into e^+e^-, similarly as the Rochester corrections in muons do.

4.4.4 Jet

The previously mentioned ID criteria for PF-jets provide efficiencies in both data and MC (derived through the tag-and-probe method) very close to one, so no SF is needed to align simulations' and observed collisions' values. However, there are other modifications that aim to improve energy measurements and correct the efficiency imbalance between data and MC with heavy-flavour jet tagging.

4.4.4.1 Energy scale and resolution

All jets undergo a complex process through which the information from the detector is exploited as much as possible to improve the precision of their energy measurement, and both the jet energy scale and its resolution from data and MC are reconciled. These corrections are applied to data, and some of them also to simulations. The CMS Collaboration follows a factorised approach that can be reviewed in detail in [177]. This process is sequential, each step attempting to correct one source of deviation between data and MC, effectively applying one SF per move. A quick summary follows of the entire method.

The first step, also called level one correction (L1), removes the energy inside the jets that comes from pileup interactions, allowing the next steps to not depend indirectly on the instantaneous luminosity of the events. These corrections are derived from a MC sample simulating QCD interactions that yield two jets in their final state (a "dijet" sample) with and without PU collisions. The differences (residues) between data and MC are derived through the random cone method with a trigger whose requirements are very general and soft (called zero-bias trigger). The L1 correction is applied to both data and MC (differently, in order to reduce the differences among them).

Secondly, and only for simulations, the response of the jets in terms of p_{T} and η is corrected. Once again, a QCD dijet sample is used to extract the corrections, comparing particle-level jets to reconstructed ones. Previously these corrections were made in two steps (level two and level three, or L2 and L3), but nowadays its only one: L2L3 (naming being kept for historical reasons).

Finally, residuals between data jets and the response-corrected simulated ones are derived and applied as corrections to the data information. This fourth step is named L2L3Residuals.

4.4.4.2 b-tagging

The use of heavy-flavoured jet tagging algorithms, such as the b-tagging one used in this thesis, has different efficiency on data and MC. SF are derived for a particular b-tagging WP that are used later to reweight the events so that the physical observables (and the possible criteria imposed upon them) related with b-tagged jets are seen with similar efficiency between data and MC. These SF are extracted separately for three sets of quark-induced jets: b-quark, c-quark, and light-quark or gluon induced jets. This is done because of the difference in the jet features due to the divergence in properties of the origin quarks (most importantly, their mass or absence of it) and how these are propagated to the subsequent generated hadrons. In addition, the b-tagging corrections also take into account the so-called "mistagging" efficiency, which is the efficiency of b-tagging with a selected WP when the selected (simulated) jet is not a b-quark induced jet, i.e. it is a jet induced by other quark flavour or a gluon. Both the SFs and the efficiencies are usually obtained depending on kinematic and geometric properties of the jets, such as its p_T and η. The efficiencies (and afterwards, the SF) are estimated in both data and MC in dedicated phase space regions enriched in the aimed jets: a detailed view of this process is explained in [178].

The obtained SF are afterwards used in the analyses by reinterpreting the efficiencies as estimations of *probabilities* of tagging as b-quark induced jet any kind of jet (tagging when the jet do come from a b quark, and mistagging in the other cases). This allows to model the probability of having $n_{\text{tagged}}^{\text{MC}}$ b-tagged jets in a simulated event, and $n_{\text{not tagged}}^{\text{MC}}$ others that are not tagged as

$$P_{\text{MC}}(n_{\text{tagged}}^{\text{MC}}, n_{\text{not tagged}}^{\text{MC}}) = \prod_{i=1}^{n_{\text{tagged}}^{\text{MC}}} \left(\varepsilon^{\text{MC}}(j_i)\right) \cdot \prod_{k=1}^{n_{\text{not tagged}}^{\text{MC}}} \left(1 - \varepsilon^{\text{MC}}(j_k)\right), \qquad (4.5)$$

where $\{j_i\}_{i=1}^{n_{\text{tagged}}^{\text{MC}} + n_{\text{not tagged}}^{\text{MC}}}$ is the total collection of reconstructed and identified jets in the event, and $\varepsilon^{\text{MC}}(j)$ is the probability (or efficiency) to be b-tagged with the chosen WP of the j jet in MC. As the SF verify by construction, for a particular jet j, $\varepsilon^{\text{data}}(j) = \text{SF}(j) \cdot \varepsilon^{\text{MC}}(j)$, the equivalent probability for data is trivially:

$$P_{\text{data}}(n_{\text{tagged}}^{\text{data}}, n_{\text{not tagged}}^{\text{data}}) = \prod_{i=1}^{n_{\text{tagged}}^{\text{data}}} \left(\text{SF}(j_i)\varepsilon^{\text{MC}}(j_i)\right) \cdot \prod_{k=1}^{n_{\text{not tagged}}^{\text{data}}} \left(1 - \text{SF}(j_k)\varepsilon^{\text{MC}}(j_k)\right). \qquad (4.6)$$

With these two values for one event, we can derive the extra weight to be applied for it simply as $w = \frac{P_{\text{data}}}{P_{\text{MC}}}$. These final MC efficiencies, $\varepsilon^{\text{MC}}(j)$, used to estimate the probabilities, are derived in a phase space region close to the signal one. The used b-tagging SF are displayed in Fig. 4.9, depending on the p_T of the jets, derived directly from signal and background events the MC simulations used in the analysis. The reweighing for simulations are extracted with the efficiencies from Figs. 4.10, 4.11, and 4.12.

(a) 2016. (b) 2017. (c) 2018.
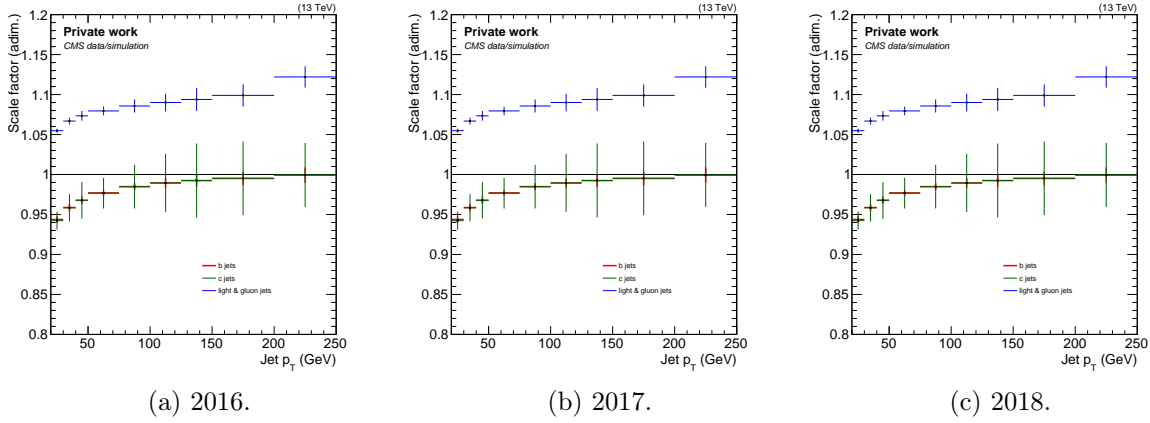
Figure 4.9: b-tagging SF used by simulation events in the analysis for the different data-taking years.

(a) 2016. (b) 2017. (c) 2018.
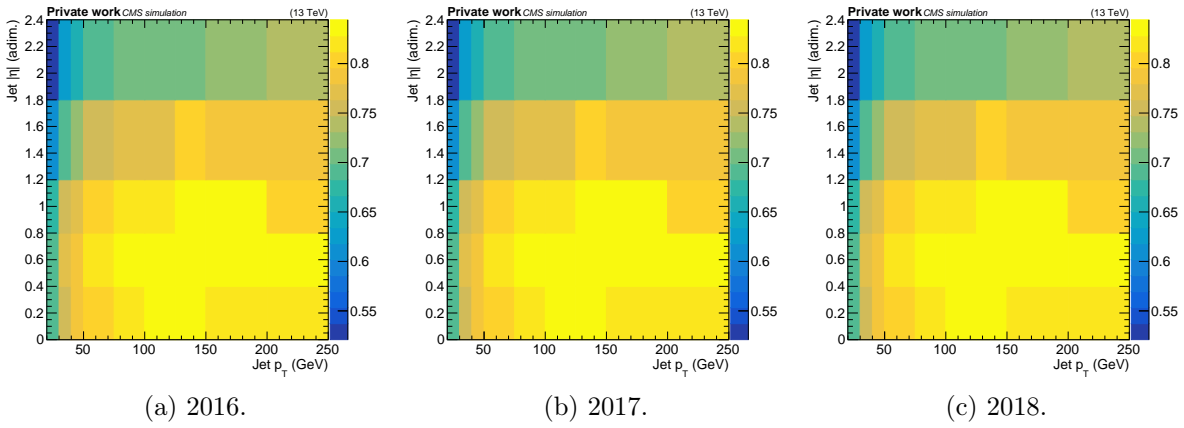
Figure 4.10: b-tagging efficiencies for b-quark originated jets of the different data-taking years.

(a) 2016. (b) 2017. (c) 2018.
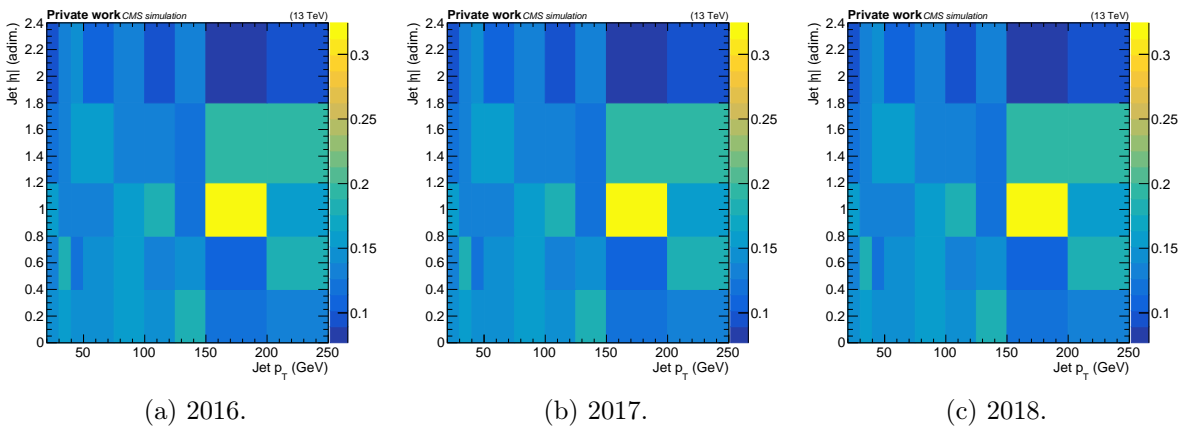
Figure 4.11: b-tagging efficiencies for c-quark originated jets of the different data-taking years.

87

(a) 2016. (b) 2017. (c) 2018.
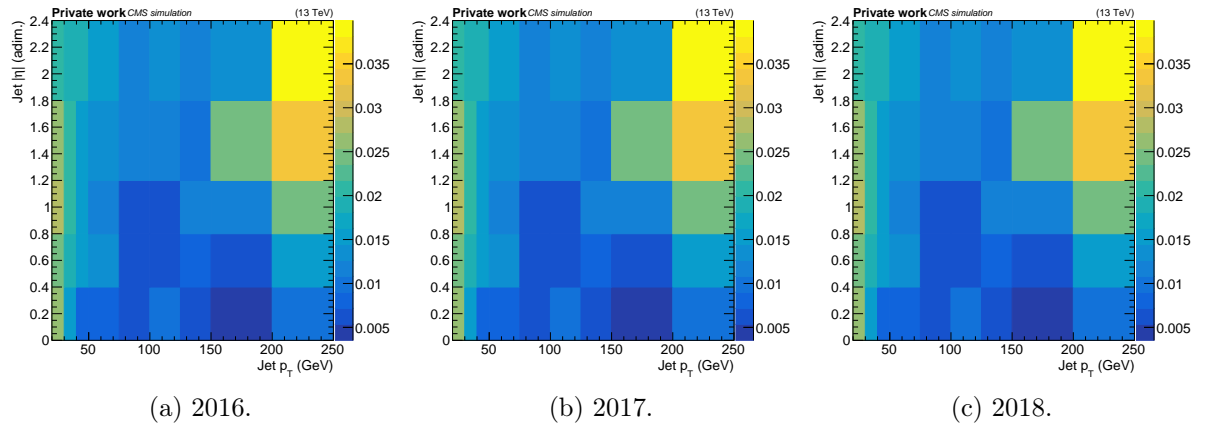
Figure 4.12: b-tagging efficiencies for light-flavour quark and gluon-originated jets of the different data-taking years.

4.4.5 Others

In 2016 and 2017, the ECAL endcap component (EE) L1 trigger showed a systematic deviation in the timing estimation of its inputs. This originated an effect in the recorded data called **prefiring**, in which events could have been stored with some of their particles assigned to previous collisions. The origin of the issue is the progressive darkening due to the ageing the ECAL crystals suffer. In 2018, this effect was taken into account in the estimation of trigger primitives.

Before that, events with jets of large pseudorapidity ($2 < |\eta| < 3$) and p_T (> 100 GeV) could be affected by the issue, effectively appearing as a trigger inefficiency but only in data. A reweighting was designed for MC to remove the difference in trigger efficiency because of this problem. To obtain it, the probability of the event to not "prefire" is calculated, depending on the reconstructed jets and photons.

Another experimental issue affected some data collected in 2018, where a fraction of the endcap component of the HCAL (HE), stopped working because of a power shut prompted by a fire alarm. This reduces the information for the corresponding geometrical area ($-3.0 < \eta < -1.3$, $-1.57 < \varphi < -0.87$) that those events had, indirectly affecting to the resolution of hadrons. There are estimations of the effects this issue might have had on the information of reconstructed jets that can be estimated using MC simulations, thus allowing analysers to evaluate the consequences in their studies.

5 Uncertainty estimation and propagation

A s in almost all situations an observable's value without its uncertainty means nothing, obtaining the latter is essential for any scientific measurement. And, given the numerous aspects of the experimental work in collider HEP listed until now, one could imagine that its estimation can be *slightly* complicated. For a general measurement at the LHC detectors, such as the cross section of a process, the following uncertainties are considered.

Statistical sources of uncertainty are the most straightforward to think of. They are directly related with the amount of data we have collected, and they can be reduced by simply taking more data. Although the $138\,\mathrm{fb}^{-1}$ integrated luminosity of data is a very large dataset, we take into account its inherent statistical uncertainty in the analysis.

Systematical sources, on the contrary, not necessarily. They are one of the most important parts of any particle physics analysis, as they tend to reflect the knowledge the physicists have of the collider, detector, and simulations they work with. In a common HEP measurement, there are three large groups to be considered. Experimental sources contain all the uncertainties associated to the corrections applied to data and MC due to the detector performance, such as PU, lepton, trigger, and jet corrections, luminosity calibrations[1], and others (such as the effect of prefiring). Modelling sources consider the uncertainties that affect the generation of the MC samples, like the chosen μ_R and μ_F scales in the ME, the tuning of PS parameters, the α_S and PDFs, and other process-dependent features. The third group will be made of the uncertainties of the cross sections used to normalise the MC. Finally, the simulation sample size, although of statistical origin, is a systematic uncertainty outside of the three groups.

Physical analyses usually need to consider specific sources because of the strategy, process, or measurement type. A clear example are differential measurements, where as we mentioned before, an unfolding must be done, commonly with their associated extra uncertainties. Or the case of maximum likelihood fits, where the estimation of templates for shape uncertainties and the incorporation of the normalisation ones must be carefully done to not add artificial bias due to e.g. low statistics in the MC used to estimate the templates.

In addition, the propagation of all sources must be carefully taken in order to not loose or ignore correlations that exist among e.g. processes in MC samples, or between data-taking periods: an example being the evolution of the very CMS detector through the Run 2, affecting the

---

[1]Results in HEP usually show the luminosity-related uncertainties separated from the systematic sources, given that its origin is more related with the accelerator/collider conditions than the detector that records the data.

experimental corrections applied and their correlations among time periods.

In the following sections the precise sources of uncertainty considered in the analysis are detailed. The correlation of the effects of any source is always total (100% correlation) or none (0%).

5.1 Experimental sources

Luminosity The uncertainty in the integrated luminosity, needed to normalise the MC to data, is of 1.2, 2.3 and 2.5% for the $36.3\,\mathrm{fb}^{-1}$ of 2016 [179], the $41.5\,\mathrm{fb}^{-1}$ of 2017 [180], and the $59.7\,\mathrm{fb}^{-1}$ of 2018 [181], respectively. These total uncertainties per measurement can be combined to achieve a total 1.6% over the $138\,\mathrm{fb}^{-1}$ of the three years. However, these uncertainties are not included in the measurements as-is: they are separated into various components so that proper correlations can be established across the years of data taking.

These parts can be reduced to the following. Three uncorrelated contributions of 1.0, 2.0, and 1.5% for 2016-2018 with components of e.g. strong statistical origin, such as background estimation. Then, two correlated components that contain effects that spread over more than one year, e.g. the uncertainty due to the potential bias of the factorisation between the x and y assumption (needed for the van der Meer scans). The first component applies 0.6, 0.9, and 2.0% for 2016-2018 and the second only applies to the last two years of the period, with 0.6% and 0.2% correlated uncertainty for 2017 and 2018.

Pileup The pileup reweighing used to correct the average PU distribution in simulations to that of data is affected by the uncertainty on the proton-proton inelastic collision cross section measurement, that was of 4.6% as mentioned before [175]. This value is propagated to the measurements in this analysis.

Trigger corrections The uncertainty sources that affect the trigger corrections detailed in Sect. 4.4 are estimated in the analysis' measurements by varying the SF by its uncertainties. These are separated into two components: statistical ones, uncorrelated per year, and systematic (due to e.g. estimating the effect of choosing the used unrelated set of triggers), correlated across years. Globally, the uncertainties are round 0.7%.

Lepton corrections As in the case of the trigger ones, the lepton corrections uncertainties are mostly estimated through the same process (varying the SFs by their uncertainties). For muons, the components are separated as the triggers' into statistical (uncorrelated) and systematic (correlated across years) origins, as well as between ID and isolation sources. An extra 0.5% of uncertainty is added in quadrature to the muon isolation systematic component to account for the extrapolation from the phase space in which the isolation SFs are measured and the phase space for the analysis. Electron reconstruction and identification

sources are considered correlated across years. Globally, they are roughly 1.5% (ID and isolation), with dependence on p_T and η.

Uncertainties coming from the lepton energy corrections applied have very little effect. Nevertheless, muon Rochester corrections uncertainties are taken into account in the analysis. They can be parameterised to the muon p_T final estimation, and are estimated in the analysis by varying this observable by them.

Jet corrections Energy corrections to jets explained in Sect. 4.4.4.1 are subject to many different sources of uncertainty. As in previous cases, we consider a grouping of these individual components that takes into account the correlations that can be established across years and effects. Here follow a brief description of the reduced 21 groups considered, mostly based in the detector sub-regions of interest for jet-energy corrections: the forward hadron calorimeter HF lies in $|\eta| > 3.0$, the ECAL without tracker cover EC2 is in $3.0 < |\eta| < 2.5$. The ECAL that overlaps with tracker EC1 covers $1.3 < |\eta| < 2.5$ and the rest, $|\eta| < 1.3$, is assigned to the barrel BB: however, these two are usually taken together as BBEC1.

- Statistical uncertainties are uncorrelated per year and per effect or origin of the corrections. One component for HF, EC2 and BBEC1 and per year is considered. Another three (one again per year) are used to take into account the statistical effects from the absolute scale uncertainties, the statistical effect on minimum-squares fit to the effect on energy of ISR and FSR against η, and the variations between the sub-periods (eras) of data taking, as part of the jet corrections depend on them and not on the year.

- A correlated-across-years source dubbed "Absolute" groups many systematic effects. They are the absolute energy scale, the ISR/FSR corrections mentioned before, the effect of hadron fragmentation (based on the `Pythia` and `Herwig` generators), the PU correction to match between data and simulations the μ distribution, and the offset due to PU in the jet p_T. To these, a dependence on η of the ISR/FSR corrections must be added.

- Another term named "FlavorQCD", correlated across years, contains the uncertainties on the differences between the responses of different-flavour quark (and gluon) -induced jets, estimated through `Pythia` and `Herwig`.

- The "RelativeBal" uncertainty are the difference between two methods used to adjust the energy against the p_T of jets, both based in a reference particle (e.g. a Z boson). One takes advantage of the missing transverse momentum (called missing projection fraction or MPF), and the other on the p_T of the reference object. This term is correlated across years.

- Another three entries correspond to the across-years uncorrelated effects of the difference in the relative residuals when the MC used is changed between different simula-

tions (e.g. between Z+jets and the *dijet* processes).

- Finally three sources, one per region (HF, EC2, and BBEC1), include the remaining systematic effects. These include the differences due to PU on the p_T in each subregion. In addition, relative corrections on η due to the JER SFs on EC1, EC2, and HF are added in this group.

Apart from the energy corrections to jet, uncertainties from the b-tagging weights used in the analysis are taken into account. They are estimated from the SF uncertainties, obtaining thus variated weights for each variation. The effects are separated in the heavy-flavour SFs (i.e. those derived for b and c-tagging) and light-flavour and gluons ones. In addition, for both, four sources are obtained: three terms uncorrelated (corresponding to the statistically independent uncertainty per year and thus one for 2016, 2017, and 2018) and one correlated for the three years (corresponding to the systematic part of the uncertainty). A total of eight b-tagging related uncertainty sources are thus incorporated in the measurements. Overall, they amount roughly to 2% for b-tagged jets and 10% for those misidentified.

Other corrections A 20% uncertainty in the corrections from the prefiring issue described in Sect. 4.4.5 is propagated into the analysis, by varying the reweighing accordingly.

5.2 Modelling sources

Limited size of MC The statistical uncertainty due to the number of generated events for simulations is taken into account as a systematic uncertainty.

Matrix element Uncertainties in μ_R and μ_F scales are estimated varying these by factors of 2 and $\frac{1}{2}$ simultaneously, and separatedly. All possible combinations are considered, except those where opposite variations of the scales happen (e.g. $2 \cdot \mu_R$, $\frac{1}{2}\mu_F$). Then, an envelope of all these variations for each observable is used as the uncertainty estimation. This source of uncertainty is considered correlated across years and separately for the tW and $t\bar{t}$ processes: these two account for the vast majority of the events in the signal regions defined in the following subsections.

PDF and α_S As mentioned in Sect. 3.3 the samples of tW and $t\bar{t}$ use the NNPDF3.1 dataset. Its uncertainty is estimated through the use of a diagonalised Hessian matrix with 100 independent entries, that allow to obtain the same number of variations, each of them accounting for an independent uncertainty. Thus, the uncertainty is obtained by summing the differences to the nominal values quadratically. To this uncertainty, the α_S's one is added, also quadratically, obtained directly from its uncertainty propagated as a reweighing. This uncertainty is correlated between $t\bar{t}$ and tW events and also across years.

Parton shower Various effects are considered that concern the PS.

- Underlying event: the effect of the tuning described in Sect. 3.2 on the underlying event modelling is assessed by varying each fitted parameter by its uncertainty for $t\bar{t}$ and tW events [166, 182]. As the tune is common for both processes, this source is correlated between them, as well as across years.

- ME/PS matching: this uncertainty is estimated through the value of the h_{damp} parameter of the POWHEG ME, $h_{damp} = \left(1.379^{+0.926}_{-0.505}\right) \cdot m_t$ [166, 183]. This setting, used through a damping function $\frac{h_{damp}^2}{(p_{\mathrm{T}}^2 + h_{damp}^2)}$, controls the merging of QCD radiation between ME and PS. It is estimated for $t\bar{t}$ simulations through the use of variated samples generated with this parameter modified: the difference with respect to the nominal values is taken as uncertainty. It is correlated across years.

- ISR and FSR: the energy scale used in the PS to simulate both ISR and FSR is modified by factors 2 and $\frac{1}{2}$ and the difference with respect to the nominal values is considered its associated uncertainty. The motivation for these factors come from the uncertainties in the PS tuning [182]. The uncertainty is estimated for $t\bar{t}$ and tW events, and it is correlated across years. For FSR, the source is correlated also between tW and $t\bar{t}$. For ISR, because the PS approximates higher-order ME contributions, they are taken separately for the two processes.

- Colour reconnection (CR): there are various algorithms that model the QCD charge links among coloured particles. To estimate the uncertainty in this simulation feature, various algorithms as described in Sect. 3.3.1 are used and alternative samples with each of them are generated for $t\bar{t}$ and tW [168]. Then, the variations of each of the three models (with ERD activated or ERDon, QCD-based or CR1, and gluon-move based or CR2) with the nominal are taken as uncertainties.

Top quark mass This value, nominally set for all MC samples as 172.5 GeV, is varied in alternative samples of the tW and $t\bar{t}$ processes by 1 GeV, that corresponds to twice the uncertainty in the measured value of m_t [184]. The difference with the nominal values is taken as uncertainty and is correlated across years and between tW and $t\bar{t}$ events.

Top quark p_{T} Differential measurements of the $t\bar{t}$ cross section depending on the p_{T} of the top quark show tensions in its modelling with POWHEG +Pythia 8 [185, 186, 187], in which the predictions are lower in average that the observed data. An uncertainty source is added to take into account this mismodelling, estimated from a reweighing extracted between generator predictions and differential measurements at parton level. The difference of this variation with the nominal values is taken as uncertainty in $t\bar{t}$ events, being correlated across years.

DR/DS different approaches The artificial modification of the tW process nominal samples is covered by an uncertainty estimated by the difference between the DR and DS predictions

for the same ME and PS generators. It is correlated across years.

5.3 Background normalisation sources

The uncertainty in the background normalisation is set per group of processes. A 4% [63] uncertainty is taken for the $t\bar{t}$ group (i.e. $t\bar{t}$ dileptonic).

For DY, 10% of uncertainty is used, roughly the double of the latest value observed for the same energy at the centre of mass measurements, that have a precision of roughly of 5% [188].
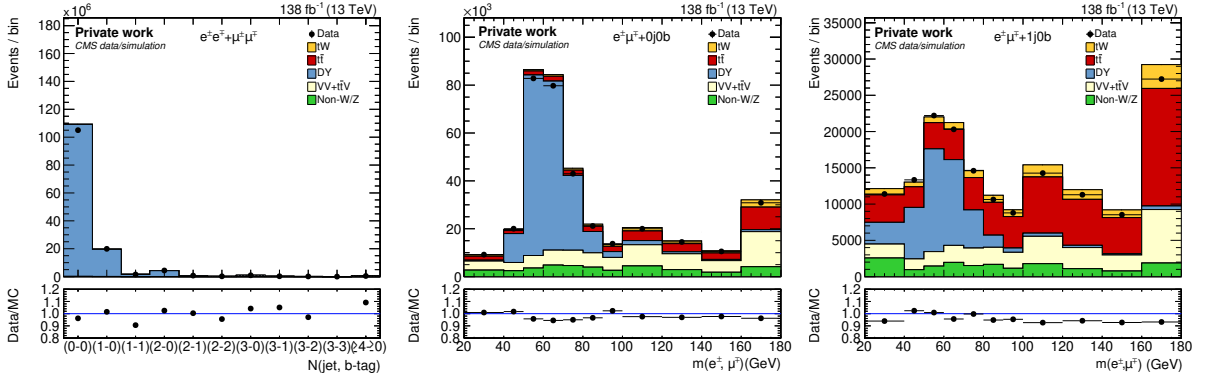
Figure 5.1: Data/MC comparison histograms in DY-enriched regions of the phase space with dileptonic requirements: the number of jet and b-tagged ones in same flavour channels (top left), and the invariant mass of the system of the two leading leptons in the $e^{\pm}\mu^{\mp}$ channel when vetoing the presence of leptons (top right) and when requiring one that is not tagged (bottom). Vertical bars on the points show the statistical uncertainty in the data. The lower panels show the ratio of the data to the sum of MC. No uncertainty is shown for the simulations.

These values are established after checking that DY-enriched regions of the phase space, close to the signal ones defined in Sect. 6.1, have a good data/simulations agreement, well under that value. These comparisons are shown in Fig. 5.1: events in those plots, filtered in both data and MC from large imbalance in p_T^{miss}, use the corrections and object ID from the previous sections. A simple selection is done, requiring at least two identified leptons (either muons or electrons), where the two leading ones in terms of p_T must be of opposite charge and the leading one at least 25 GeV of p_T. To remove contamination from low-energy resonances, the minimum invariant mass of all lepton-pair systems must be at least 20 GeV. Some comparisons are done then by requiring those leading leptons to be of same flavour (i.e. the $\mu^-\mu^+$ and e^-e^+ production *channels*), or of opposite one (the $e^{\pm}\mu^{\mp}$ channel). In the latter case, two data/MC comparisons are shown, the first by requiring no identified jets in any event, and the second one demanding only one jet that must not be b-tagged.

Thanks to this selection, we obtain in the same flavour comparison a very pure region in the DY process, in the histogram of the combined distribution of the number of jets per event, and the number of those that are b-tagged. In it, clearly almost all events agree within a 10% variation.

In other many bins (particularly, in those with larger number of events), the agreement could be established under 5% and only in a couple of them it is 10% or larger, but with also much smaller statistics. Other distributions checked with this selection also showed similar agreement levels (under 10% in almost all cases, mostly under 5%).

In even closer phase-space subregions to our signal one (mentioned in Sect. 6.1), but with larger contributions from other processes, the other two plots show the reconstructed invariant mass from the two leading identified leptons. The DY contribution here appears focused around $50 - 70$ GeV, with the form of a peak whose origin is the actual Z boson resonance, but displaced because of the event selection. As we are in the $e^{\pm}\mu^{\mp}$ channel, collecting events from $\mu^{-}\mu^{+}$ or $e^{-}e^{+}$ decays is complicated due to the good reconstruction and identification efficiencies of both lepton flavours. However, a $\tau^{-}\tau^{+}$ decay where each $\tau^{\pm}$ decays to one electron and one muon can explain this shifted peak, where part of the momentum is "lost" in the painted histograms in the form of neutrinos. Around this moved resonance, the agreement lies also well under 10%.

Finally, for the Non-W/Z and VV+t$\bar{t}$V processes a 50% uncertainty is set. In Non-W/Z, the chosen normalisation uncertainty, higher than that of the latest measurements of its process, allows to cover for the uncertainty in the modelling of misidentified leptons originated from jets. Both groups have a low presence in the yields of Tab. 6.1, with less than 2% (combining both) of the total expected events.

6 Inclusive measurement

W ITH the elements depicted in the previous chapters, we are ready to make the first observation of this document: the inclusive cross section. To do so, we will need first to define a event selection, then detail how to extract the signal, before giving the result itself. The following sections contain those three steps.

6.1 Event selection

Data and MC events are filtered to a signal region in each measurement: inclusive and differential. The exact definition is different, but they are very similar. As the analysis aims to measure in the dileptonic channel, evens must have at least two identified leptons. To avoid background contamination, we focus ourselves on the $e^\pm\mu^\mp$ channel, demanding that the two leading leptons (in terms of p_T) are an electron and a muon of opposite charge. That of those with larger p_T should have $p_\mathrm{T} > 25$ GeV. Also, events must not have any lepton-pair system with invariant mass larger than 20 GeV to avoid contamination from low-energy resonances. Any event with an abnormal presence of $p_\mathrm{T}^\mathrm{miss}$ is filtered out.

These requirements define what we call our baseline selection, and allow us to further categorise events depending on the presence of jets and the number of those that are b-tagged. As shown in Fig. 6.1, this grouping of events is very useful to differentiate the process we want to measure, in particular to largely suppress all backgrounds except $t\bar{t}$ in the at least one jet and at least one b-tagged jet bins. Events that do not fulfill that criterion show larger presence of mostly DY backgrounds.

The largest signal-to-background ratio is present in the 1j1b (one jet, and one b-tagged jet) bin of this observable, with roughly 19%. Even there, there is five times more background that signal, mostly coming from top-antitop quark pair production.

To measure the inclusive cross section of tW, we build up on the baseline selection and use the 1j1b, 2j1b, and 2j2b bins of the number of jets and of those b-tagged distribution from Fig. 6.1. The first of them has the largest relative content in signal of any of them, as it is the closest definition to what we could expect from tW's final state. The 2j1b bin contains also a non-negligible contribution of signal, while 2j2b has almost none and it is used to help keep control over the overwhelming $t\bar{t}$ contribution in the signal extraction process. The predicted event content per processes-group and the collected data in each of them are shown in Table 6.1.
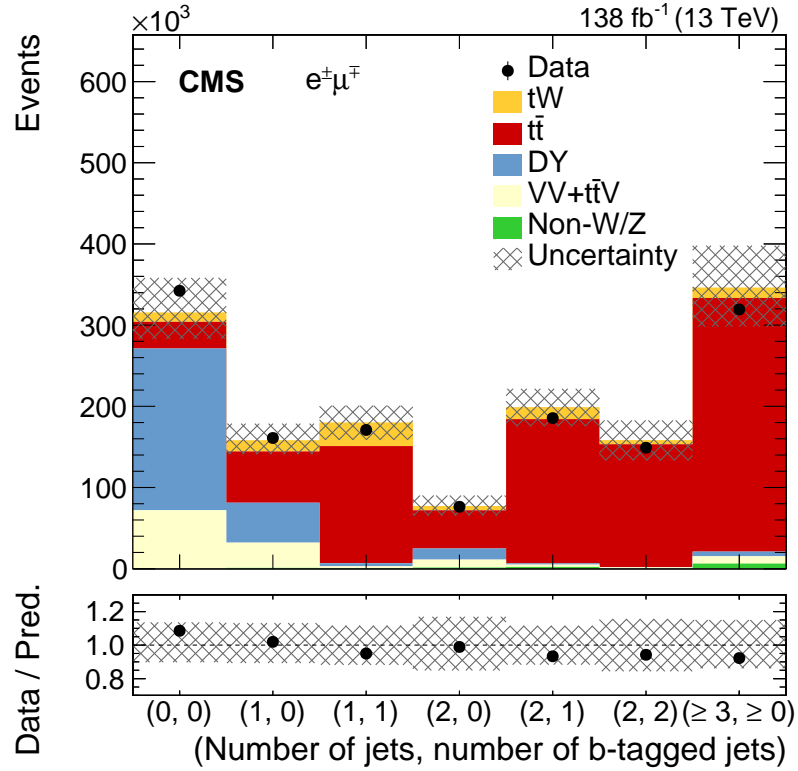
Figure 6.1: Data/MC comparison histogram showing the number of jet and b-tagged ones with the baseline selection described in the text. Vertical bars on the points show the statistical uncertainty in the data, and the hatched band the total uncertainty in the simulations. The lower panels show the ratio of the data to the sum of MC.

Process	1j1b	2j1b	2j2b
tW	$29\,000 \pm 2\,000$	$14\,500 \pm 1\,300$	$4\,800 \pm 600$
t$\bar{\text{t}}$	$140\,000 \pm 20\,000$	$180\,000 \pm 20\,000$	$150\,000 \pm 20\,000$
DY	$3\,600 \pm 500$	$1\,500 \pm 200$	210 ± 60
VV $+$ t$\bar{\text{t}}$V	$2\,400 \pm 1\,000$	$2\,800 \pm 1\,200$	$1\,400 \pm 600$
Non-W/Z	700 ± 300	$2\,200 \pm 900$	270 ± 120
Total	$180\,000 \pm 20\,000$	$199\,000 \pm 23\,000$	$160\,000 \pm 20\,000$
Data	$170\,900 \pm 400$	$185\,400 \pm 400$	$148\,900 \pm 400$

Table 6.1: Total number of events observed in data and the number of signal and background events expected from simulation in the 1j1b, 2j1b, and 2j2b regions.

6.2 Signal extraction

The signal extraction step can be achieved in different ways. The most direct one would be to just subtract the contribution estimation of background processes in the signal region to data, and given that we would have checked that we can reasonably model the data with our MC predictions, we would then reasonably assume that the result must be the signal. However, other approaches that exploit our statistical knowledge of the data can be pursued to potentially obtain better results (in terms of signal contribution estimation). One of them is what we call a **maximum likelihood fit**, that is used in this analysis.

The likelihood function, denoted as $\mathcal{L}$, is an application that parametrises the feasibility (or plausibility) of a given model, or set of parameters, according to some observation: the higher the values of the likelihood for some given arguments, more likely that model or parameters have yielded to the observed data. Thus, a likelihood for a given variable or distribution can be constructed depending on our observations, and the statistical knowledge of our data-taking processes. This could be the case of e.g. the p_{T} of one final-state jet. Afterwards, using as inputs our MC simulations into the likelihood, we can use a numerical procedure, a fit, in which beginning from those inputs, we maximise the likelihood function's value. In the steps of this procedure, the parameters of the model, such as e.g. the signal process contribution, vary in an usually vast parameter hyperspace. At the end, when this maximisation is done, the amount of events corresponding to the signal process can be extracted.

To perform this technique we must construct a likelihood function. In the maximum likelihood fits used in this thesis, we construct likelihoods based upon binned histograms of our data, using MC simulations to get an initial modelling of the observable (a "template" of it) for the different processes, including the signal one. Afterwards, the maximisation of the likelihood is done by minimising the $-\log(\mathcal{L})$ (commonly called log–likelihood), which is a better computational approach. The likelihood is constructed beginning with these foundations:

$$\mathcal{L}\left(\vec{r}, \vec{s}(\vec{\theta}), \vec{b}(\vec{\theta}), \vec{\theta}\right) := \prod_{i=1}^{N_{\mathrm{bins}}} \mathrm{Pois}\left(n_i | r_i, s_i(\vec{\theta}), b_i(\vec{\theta})\right) \cdot \prod_{j=1}^{N_{\mathrm{uncs.}}} e^{\frac{\theta_j^2}{2}}. \tag{6.1}$$

Here, $\vec{n} = (n_1, \ldots, n_{N_{\mathrm{bins}}})$ represent the observed values (the data) in each bin i. The arguments of the likelihood are each r_i, which is a parameter that scales the amount of signal in the bin i, s_i, whereas b_i represents the amount of background in it. The vector $\vec{\theta}$ is the last input of this function and it is a collection of several parameters which are in the likelihood to represent other information regarding our analysis, essentially (and for this document's fits, only) uncertainties [189, 190]. The $\mathrm{Pois}(n|\nu)$ represents a Poisson probability density function, that is defined as $\mathrm{Pois}(n|\nu) := \frac{e^{-\nu}\nu^n}{n!}$. In our case, the arguments are:

$$\text{Pois}\left(n_i | r_i, s_i(\vec{\theta}), b_i(\vec{\theta})\right) = \frac{1}{n_i!}\left(r \cdot s_i(\vec{\theta}) + b(\vec{\theta})\right)^{n_i} e^{-\left(r \cdot s_i(\vec{\theta}) + b(\vec{\theta})\right)}$$

$$s_i(\vec{\theta}) = s_i^T \prod_{k=1}^{N_{\text{norm.}}^S} q_k^{\theta_k} \prod_{m=1}^{N_{\text{shape}}^S} p_m(\theta_m) \tag{6.2}$$

$$b_i(\vec{\theta}) = b_i^T \prod_{l=1}^{N_{\text{norm.}}^B} q_l^{\theta_l} \prod_{m=1}^{N_{\text{shape}}^B} p_n(\theta_n)$$

In these expressions, s_i^T and b_i^T represent the total amount of events in the bin i of signal and background respectively, $N_{\text{norm.}}^S$ and $N_{\text{norm.}}^B$ show the amount of normalisation uncertainties (sources whose effect is only in the total amount of events of a given process, not on its distribution' shape) in signal and background (resp.) and the values $q_k^{\theta_k}$ and $q_l^{\theta_l}$ are factors that represent those uncertainties. E.g. if one of them were of the 20%, then $q = 1.20$ and thus $q^\theta = 1.20^\theta$. The factors $p_m(\theta_m)$ and $p_n(\theta_n)$ represent how each uncertainty that affects the shape of the distribution influence the total amount of signal and/or background. They are proportional to a quadratic polynomial that depends on the corresponding θ whose coefficients are determined before the algorithm begins to iterate through a vertical interpolation in each bin i taking into account the varied shapes of the distribution for each uncertainty (one time when a $+1\sigma$ uncertainty is propagated into one distribution and another when is the -1σ). In a nutshell, we can separate two groups of uncertainties that are represented by those parameters θ: the normalisation, and the shape ones, and each of them has an additional Gaussian term as appears at the end of Eq. 6.1.

These Gaussian terms, that conform the second factor of eq. (6.1), make our likelihood partly Bayesian, as it expresses a degree of belief of the probability of those θ variables. This term is used to introduce the effects of the uncertainties thanks to the θ parameters associated to them. When the minimisation algorithm iterates, new values of all parameters θ and r_i are taken and with them s_i and b_i are derived.

The parameter r_i is our main aim in this procedure, as afterwards we can just multiply it by s_i (for each bin) obtaining the signal (what, after all, we wanted). Thus it is called the **parameter of interest** (POI). The remaining parameters of the fit, although necessary in order to construct a realistic statistical model, are "secondary" and not of our interest, and thus they are called **nuisance parameters**, or more commonly, "nuisances". One particular situation appears when the uncertainty source of the limited size of the simulation samples needs to be considered. A method, called Barlow-Beeston, allows to do so by modifying the likelihood from Eq. (6.1) as described in [191].

The analysis of this thesis constructs a maximum likelihood fit over the events of the three regions mentioned in the previous chapter: 1j1b, 2j1b, and 2j2b to extract the signal. To enhance its separation power, the events are distributed depending on one different observable in each region over various bins.

Here follows the description of the variables used for each one. For the 2j2b region, as the presence of signal is very small, the p_T distribution of the subleading jet is used as distribution. This variable is sensitive to JES variations and is useful in constraining this systematic uncertainty in the fit. For the two remaining regions, as there is no single observable that gives strong discrimination between $t\bar{t}$ and tW events, two independent MVAs or multivariate analysis are used in this thesis. MVAs combine several characteristics (variables) of the events or physical objects at the same time to discriminate events. Their advantage is that although those variables by their own might not be a very good discriminator, the combination of their information can yield a more powerful separator. This is done the scan of the multidimensional space of all of them at the same time can be better, despite not being easy to understand a hyperspace of several dimensions.

A **boosted decision tree** (or BDT) [192, 193] is a clear example of an MVA algorithm that is used in this thesis. They can be understood as an evolution of the "traditional" tool of the decision trees, which are used for the aim of classification of events. The main idea behind boosting is to construct a powerful learner out of an ensemble of weak learners. In the case of a boosted decision tree, an ensemble of shallow trees is trained. The training of the trees is performed sequentially, and the training of each tree depends on the false positives of the previously trained trees. There are different boosting algorithms, though their main idea is the same: after each iteration of the algorithm, improve the classification of the events by varying the selection criteria. At the end, all BDT are able to classify events by associating each one with a value called discriminator, which encompasses all the classification of the several variables. This value usually goes from -1 to $+1$ and the idea is that the BDT is usually trained so that events that have values near $+1$ tend to be more signal–like, whereas those near -1 are more background–like.

The BDTs used in the analysis are trained and tested using a set of simulated samples that are statistically independent from the ones used in the signal extraction. They are trained to discriminate between $t\bar{t}$ and tW. 70% of them is used for the actual training, and the remaining 30% for testing. For both models, hyperparameters (number of trees, learning rate, and the maximum tree depth) are optimised comparing with the area under the curve (AUC) of the receiver operator characteristic (ROC) graph, that allows us to search for the best balance between precision and recall in the algorithm. The optimisation is done manually, contrasting AUCs obtained with a set of hyperparameter configurations.

As with any ML-based method, the possibility of overtraining (or overfitting) the model to the information provided during training exists. The overtraining in the two BDTs is assessed after the training by comparing various figures of merit evaluated on the simulated events used to train the algorithms, and those from the test set. These are the AUC of the ROC curve, and the p-values of goodness-of-fit (GOF) tests: Kolmogorov-Smirnov (KS) and χ^2.

The input variables used in both MVAs are chosen depending on how well the MC simulation models the data and on their discrimination power. To select them, several (≈ 50) candidates

have been considered for both 1j1b and 2j1b categories. They were observables from reconstructed particles or combined systems of them, sometimes considering p_T^{miss} as an extra addition, and trying to exploit differences between tW and $t\bar{t}$ that might exist. An early selection was done by removing those that clearly had bad data/MC agreement, by vetoing clear tendencies/patterns, and ratios significantly outside uncertainties or larger than $\pm 5 - 10\%$ (in the bulk of distributions, tail outliers were tolerated). Later, GOF tests were done to clearly assess the agreement, removing or modifying those variables with significant bad agreement. The two models were trained with the pruned sets of variables, and their hyperparameters optimised as mentioned before.

For the BDT in the 1j1b, the list of variables used in the training in order of importance are:

- $p_T(e^\pm, \mu^\mp, j)$: the magnitude of the transverse momentum of the dilepton + jet system.

- $C(e^\pm, \mu^\mp, j)$: centrality, which is defined as $\sin\theta$, where θ is the polar angle of the total momentum of the system (with respect to CMS' interaction point).

- $m(e^\pm, \mu^\mp, j, \vec{p}_T^{miss})$: invariant mass of the dilepton + jet + $\vec{p}_T^{miss}$ system.

- Leading loose jet p_T: if there are no loose jets, this variable is set to 0.

- Jet p_T.

- Presence of loose jets in the event: the result is either yes or no.

The order of importance is determined by counting how often each variable is used to split the decision tree nodes. The counts are weighted by the square of the separation gain achieved by the variable and by the number of events in the node. Figure 6.2 shows the data/MC agreement of the training variables, where overall good agreement is observed. An AUC of the ROC curve of ≈ 0.64 was achieved. Overtraining was not found in the model, after reviewing the chosen figures of merit.

The input variables listed in order of importance in the BDT for 2j1b are:

- $\Delta R(\ell_1, j_1)$: separation in η-φ space between the leading lepton and leading jet, where φ is the azimuthal angle.

- $\Delta R(\ell_{12}, j_{12})$: separation in η-φ space between the dilepton and dijet systems.

- Second-highest jet p_T (subleading).

As shown in Fig. 6.3, the data/MC agreement is contained in the $5 - 10\%$ band, within uncertainties. The trained BDT achieved an AUC of the ROC curve of ≈ 0.60. No overtraining is seen in this model, as e.g. differences in this value between train and test simulations are very small $(O(10^{-4}))$.

The binning of the two BDT output distributions is chosen such that each bin contains about the same number of $t\bar{t}$ events. This avoids the presence of low-statistic bins in the background
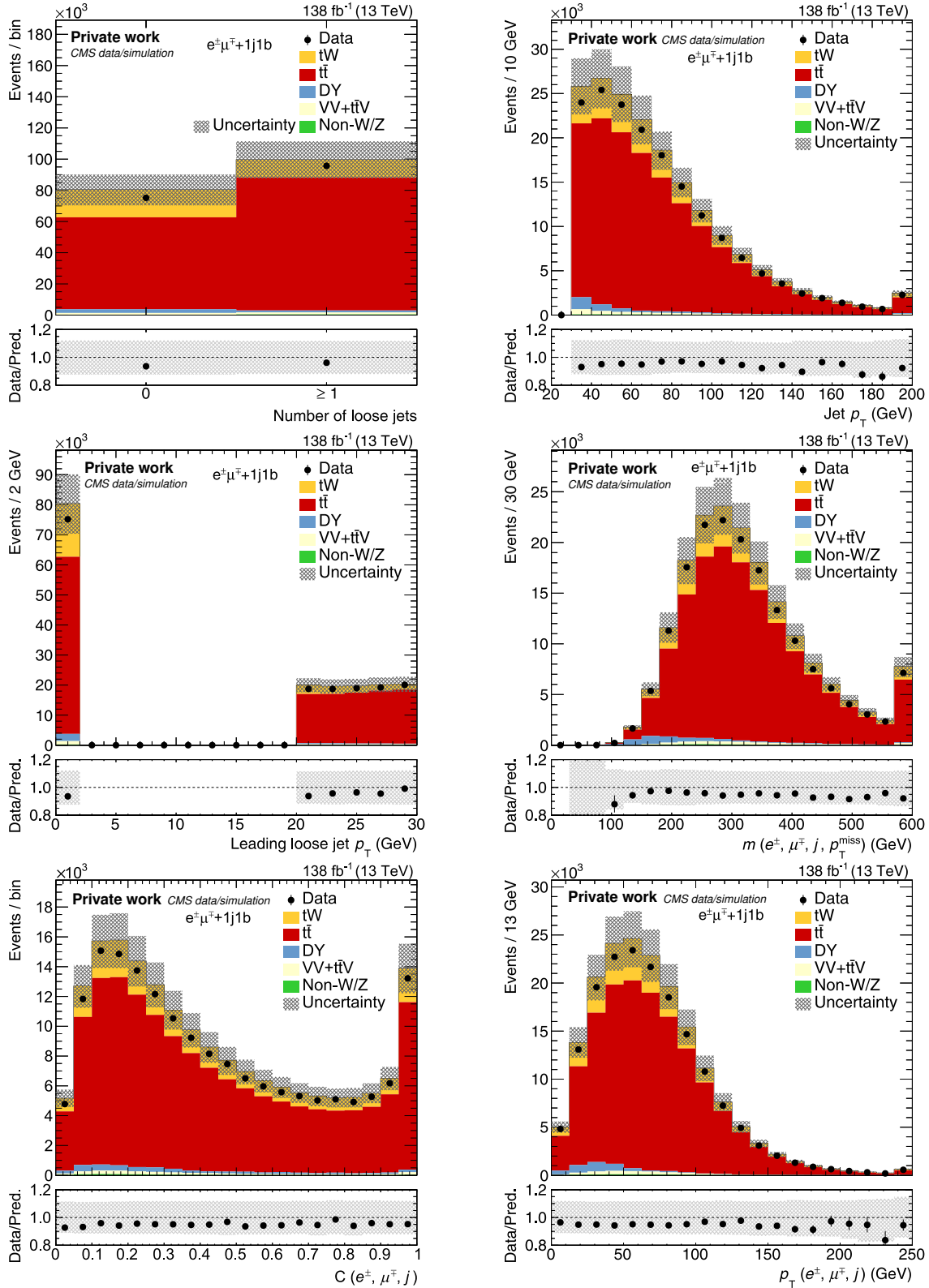
Figure 6.2: Data/MC comparison of the input variables used for the BDT trained in the 1j1b region. From top left to bottom right: presence of loose jets, jet p_T, leading loose jet p_T, $m(e^\pm, \mu^\mp, j, \vec{p}_T^{\,\mathrm{miss}})$, $C(e^\pm, \mu^\mp, j)$ and $p_T(e^\pm, \mu^\mp, j)$. Vertical bars on the points show the statistical uncertainty in the data, and the hatched band the total uncertainty in the simulations. The lower panels show the ratio of the data to the sum of MC.
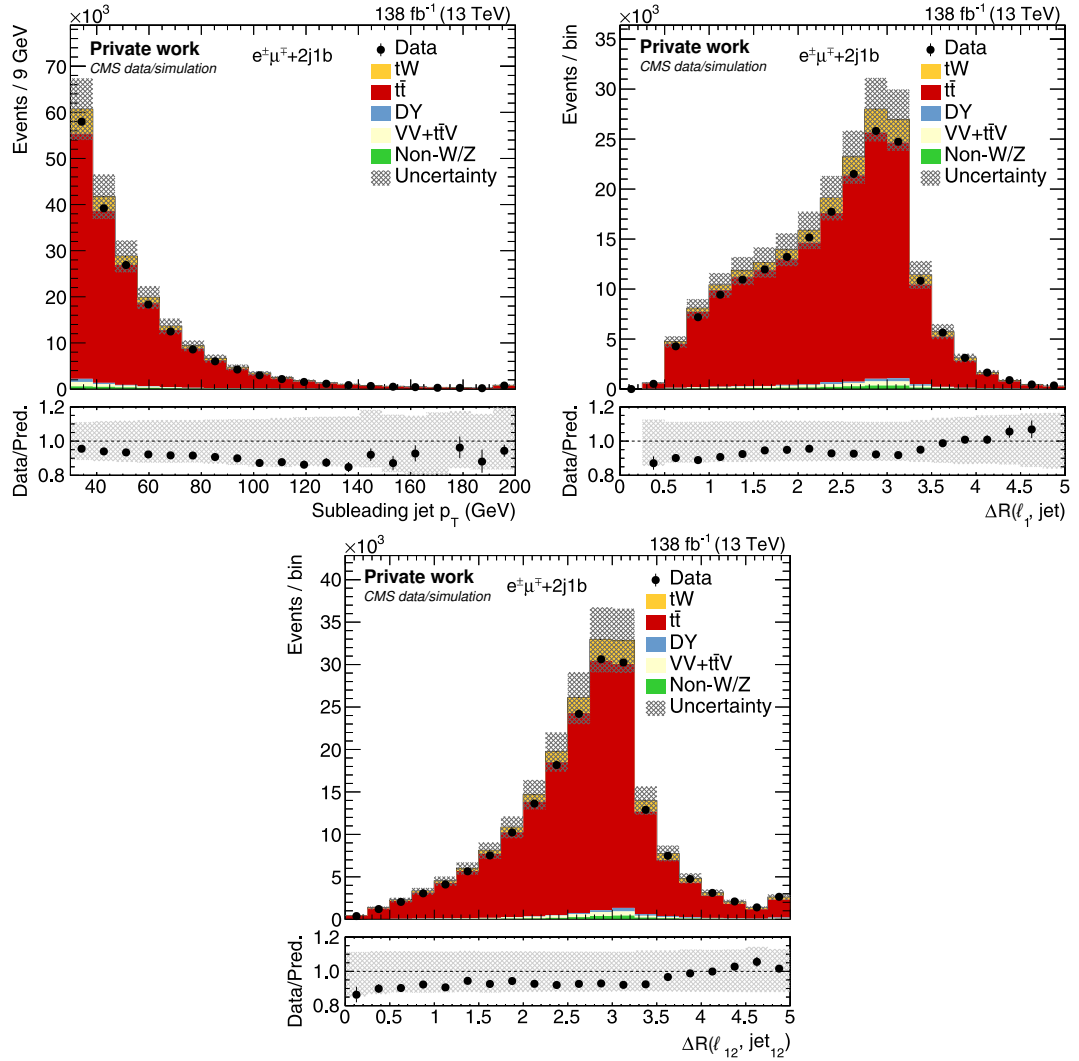
Figure 6.3: Data/MC comparison of the input variables used for the BDT trained in the 2j1b region. From left to right: subleading jet p_{T}, $\Delta R(\ell_1, \mathrm{j}_1)$, and $\Delta R(\ell_{12}, \mathrm{j}_{12})$. Vertical bars on the points show the statistical uncertainty in the data, and the hatched band the total uncertainty in the simulations. The lower panels show the ratio of the data to the sum of MC.

estimation, helping to constrain the systematic uncertainties. The number of bins in the three variables used in the signal extraction has been chosen to have as many of them (in order to exploit the information as much as possible), but without inducing large statistical fluctuations in the background and uncertainty estimations due to the lack of enough simulated events.
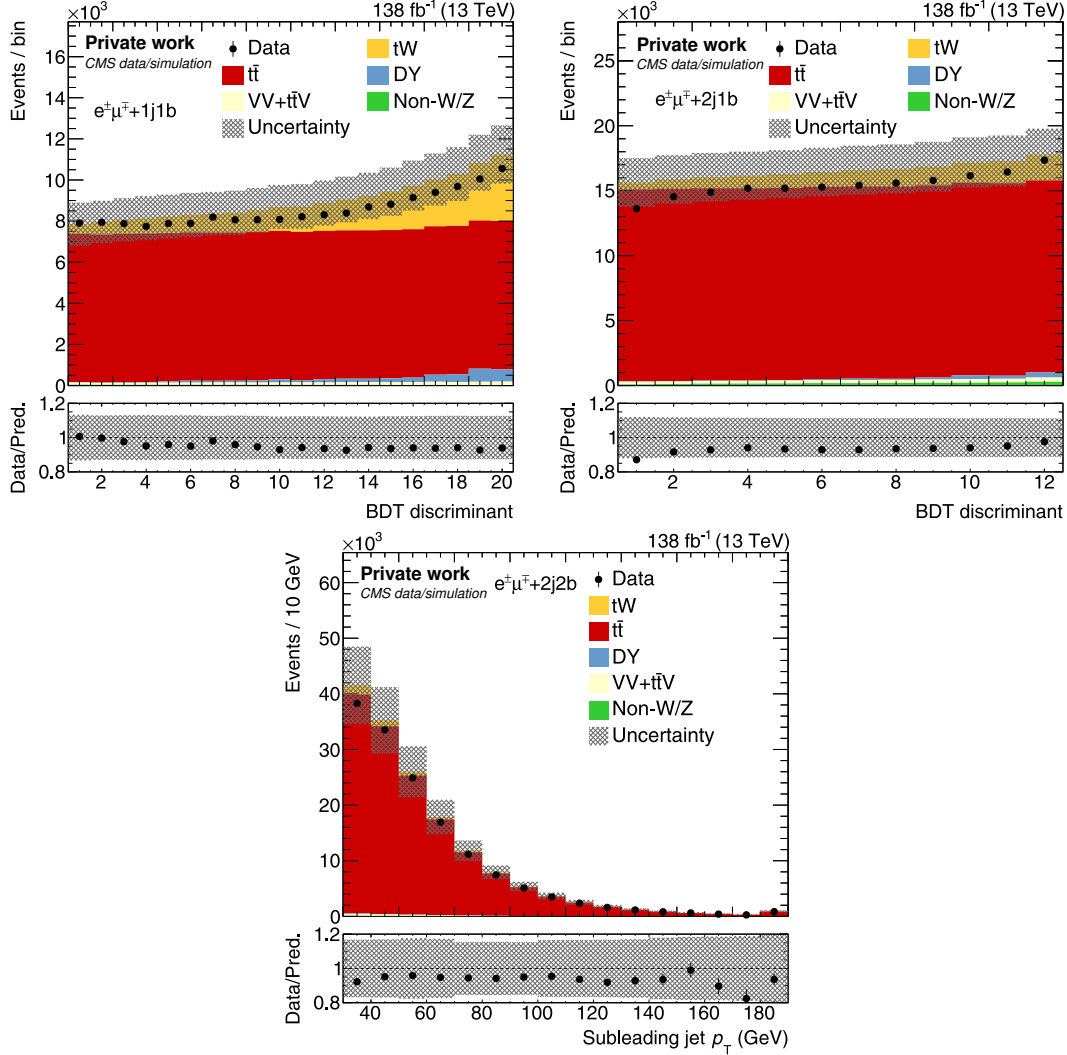
Figure 6.4: Data/MC comparison of the 1j1b (left), 2j1b (centre), and 2j2b (right) input variables of the maximum likelihood fit used in the signal extraction. Vertical bars on the points show the statistical uncertainty in the data, and the hatched band the total uncertainty in the simulations. The lower panels show the ratio of the data to the sum of MC.

Simultaneously, the maximum likelihood fit extracts the signal contribution from these three distributions, that are displayed, before the minimisation, on Fig. 6.4. The fit has only one POI (the signal strength $\mu_{tW} = \frac{\sigma_{tW}^{obs.}}{\sigma_{tW}^{theo.}}$), and is constructed as described at the beginning of this chapter. Signal and background processes templates are obtained from the simulations listed in Sect. 3.3. The uncertainty on the limited size of this MC is introduced through the Barlow-Beeston method [191]. The remaining systematic sources are considered through nuisance parameters

whose templates are estimated from the variations described in Sect 5. Statistical uncertainties enter the fit directly, in the construction of the likelihood. Correlations across data-taking periods (years) are considered by separated nuisance parameters. Correlated sources that affect only some years affect only to the simulations that model the assigned data-taking period.

Extensive work was made to avoid numerical issues with the fit's minimisation. This potential hazard was enlarged due to the large amount of observed data in the signal region (1j1b, 2j1b, and 2j2b), collected and distributed in a significant number of bins. The constrain power of the fit was substantial because of this. And, as a consequence, the minimisation result was sensitive to small fluctuations in the event estimation and in the templates used to include the shape uncertainty sources. A summary of the checks and preemptive actions taken follows:

- GOF tests assessed the original data/MC of the three input distributions, the training variables of the MVAs as mentioned before, and the combination the three regions.

- An automatised review of the fit's inputs ensured no unphysical contribution (negative content due to generator weights, or null number of events) from any MC estimation, and the removal of any low-contribution (less than 0.01 expected events) processes group (although the latter was not found needed to be applied in this fit). A similar check was done for systematic uncertainties in the fit, by checking that shape uncertainties indeed vary non-negligibly the distribution of the observable ($> 10^{-5}$ factor between two non-zero consecutive bins), and neglecting the effect of normalisation sources whose uncertainty is extremely small ($< 10^{-3}$ factor variation).

- A non-automatised review of the fit's inputs, done processes group-per-processes group (i.e. tW, $t\bar{t}$, DY, VV + $t\bar{t}$V, and non-W/Z), uncertainty source-per-uncertainty source (all correlated and uncorrelated contributions from the sources in Sect. 5, summing a total of 75 items), and variation-per-variation (two, up and down, and one when applicable), substantially improved the fit's stability and removed artifacts in its results. This detailed study helped identify small fluctuations that, due to the constrain power of the fit, affected its result by either spuriously hiding uncertainty, or creating it. Automated checks also showed that fluctuations created one-sided effects in some bins for separated processes.

These issues were solved by modifying the original uncertainty estimations, as done in other analyses of the field, through a **smoothing** of its shape. In some situations, the effect of a (originally) shape source could be smoothed by neglecting its shape effect and assuming only that in the normalisation. In others, where the shape effect was appreciable and recognisable, the ratio of the variations (separately) to the nominal expectations for the process to which the source affected was fitted to polynomials. Then, the values in the centre of the bins were taken from the resulting function *if* the fit converged and modelled satisfactory the original estimations. In addition, the new variations were symmetrised unless the original ones tended to exhibit asymmetry between them.

All these comparisons and tuning of the uncertainty estimation removed numerical issues with the minimisation and artifacts from its result. These were mainly found when reviewing the impact of the parameters of the fit. Examples of the smoothing applied to the fit's inputs are shown in Fig. 6.5.
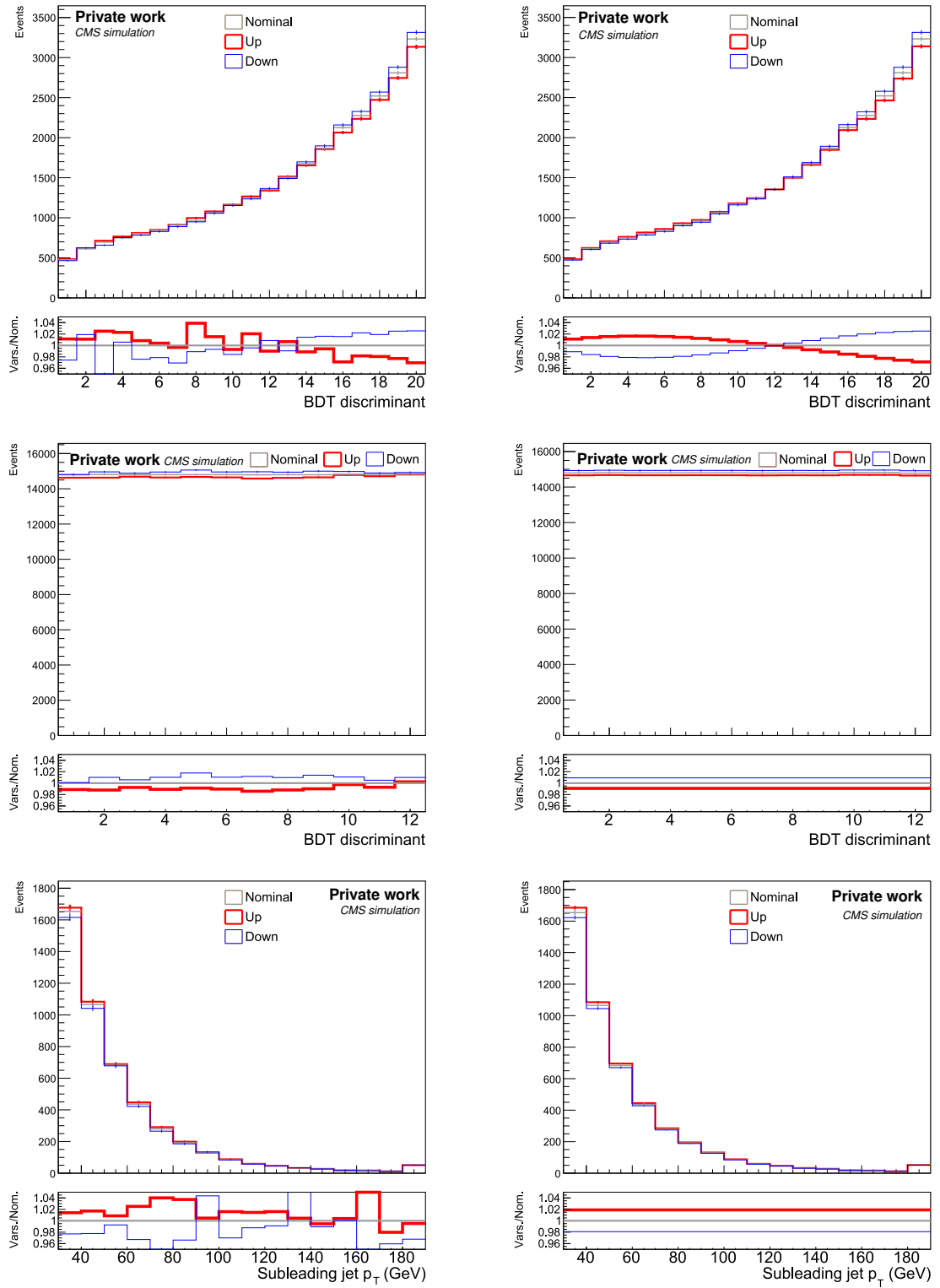
Figure 6.5: Examples of smoothing applied to uncertainty source variations on the three regions used in the inclusive measurement as described in the text. The top plots, corresponding to 1j1b, show the effect on the signal process of the JES FlavorQCD source before the smoothing (left) and after (right). The centre plots, from 2j1b, idem but of the ME/PS source on $t\bar{t}$. The bottom are from the 2j2b region and show analogously the variations of FSR on the tW process. Vertical bars show statistical uncertainty of the estimations. The lower panels show the ratio of the data to the sum of MC.

6.3 Result

The measured value for the signal strength is obtained by maximising the likelihood function with respect to all its parameters. The resulting value corresponds to an inclusive cross section of

$$\sigma_{\text{tW}}^{\text{obs.}} = 79.2 \pm 0.9(\text{stat.}) {}^{+7.7}_{-8.0}(\text{syst.}) \pm 1.2(\text{lumi.}) \text{ pb.} \tag{6.3}$$

The distributions of the BDT discriminants in the 1j1b and 2j1b regions and the subleading-jet p_{T} distribution in the 2j2b region after the fit are shown in Fig. 6.6. The observed and MC predicted event yields after the fit in the three regions are given in Table 6.2.
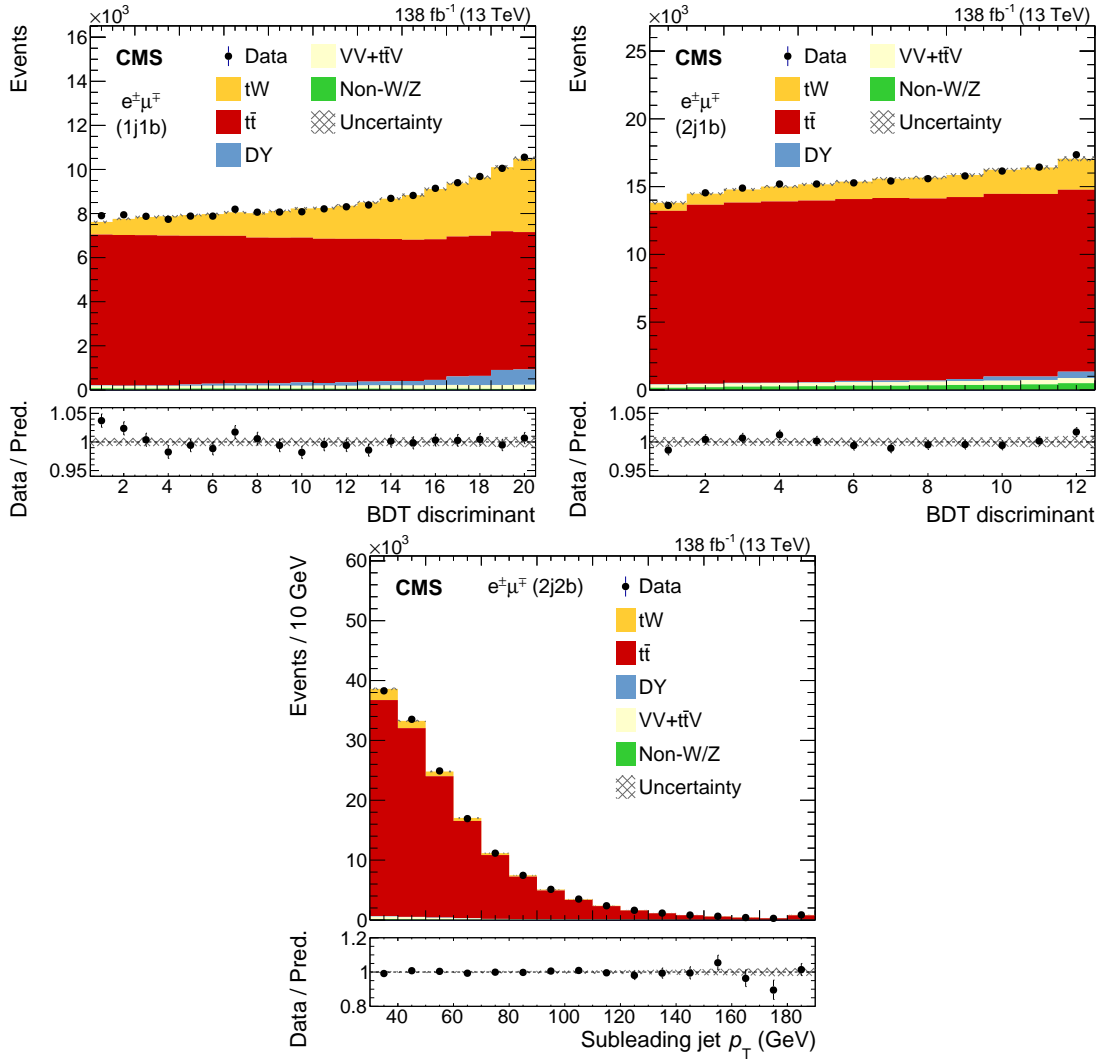
Figure 6.6: Data/MC comparison of the 1j1b (left), 2j1b (centre), and 2j2b (right) variables of the maximum likelihood fit used in the signal extraction, displaying the post-fit contributions from the different processes and its uncertainties. Vertical bars on the points show the statistical uncertainty in the data, and the hatched band the total uncertainty from the fitted expectations. The lower panels show the ratio of the data to the sum of post-fit MC.

| Process	1j1b	2j1b	2j2b
tW	$31\,600 \pm 600$	$16\,600 \pm 500$	$5\,500 \pm 200$
$t\bar{t}$	$131\,200 \pm 500$	$160\,300 \pm 600$	$141\,100 \pm 400$
DY	$3\,990 \pm 190$	$1\,630 \pm 100$	260 ± 20
$VV + t\bar{t}V$	$2\,800 \pm 300$	$3\,300 \pm 500$	$1\,700 \pm 400$
Non-W/Z	$1\,140 \pm 150$	$3\,700 \pm 700$	470 ± 120
Total	$170\,800 \pm 300$	$185\,400 \pm 400$	$149\,100 \pm 300$
Data	$170\,900 \pm 400$	$185\,400 \pm 400$	$148\,900 \pm 400$

Table 6.2: The number of observed and MC predicted events after the fit in the 1j1b, 2j1b, and 2j2b regions. The statistical uncertainties in the data and the total uncertainties in the predictions are given.

The result is clearly systematically dominated, followed at distance from the luminosity uncertainties and then the statistic source (in decreasing order). The 20 largest impacts on the signal strength and the corresponding nuisance parameters are shown in Fig. 6.7. The impact is defined as the shift $\Delta\hat{\mu}$ induced in the signal strength μ when the nuisance parameter θ is varied by ± 1 standard deviation (σ) around its best fit value. The leading uncertainties are the JES absolute correction sources, the normalisation of the non-W/Z background, the ME scales of the tW process, and the modelling of FSR for $t\bar{t}$ and tW. Other separated JES components appear in the subsequent most relevant impacts. Its large effect in the measurement comes from the importance jets have in the event selection (the $ijjb$ categories) as well as in the signal extraction. One example of this is the use of low energetic jets (loose jets) in the BDTs: they are a double edged sword. Loose jets allow to gain separation power between tW and $t\bar{t}$. However, they are subject to larger JES correction uncertainties and higher effects from FSR or PU. The non-W/Z affects essentially the 2j1b region, where its presence is most relevant: its large impact is explained by the similarity of its shape with the signal one as well as its large value (the same can be applied to the $VV + t\bar{t}V$ and DY background normalisation source, that appear both also up in the ranking).

Figure 6.7 also shows the pulls of the nuisance parameters, $(\hat{\theta} - \theta_0)/\Delta\theta$, where $\hat{\theta}$ and θ_0 are the values after and before the fit of the nuisance parameter θ, and $\Delta\theta$ its uncertainty before the fit. Various nuisance parameters, such as the b-tagging efficiency and jet energy corrections, are significantly constrained in the fit due to their effect on the jet multiplicity. The ME scales of the $t\bar{t}$ process are also constrained because of the large presence of $t\bar{t}$ events in all the regions used in the fit. These constraints help in the reduction of the overall uncertainties of the measurement.

The pulls of the $t\bar{t}$ μ_R and μ_F scales, and normalisation uncertainty, partly explain the variation in this process' yields from those of Tab. 6.1 to the values of Tab. 6.2. This reduction in the $t\bar{t}$ expected events is also in line with the latest measurements of $t\bar{t}$ at this energy from CMS [63, 194], where a lower production cross section than the NNLO+NNLL prediction was observed. The same pattern, although with less precision in the observations, is seen in the early Run 3 measurements from ATLAS [65] and CMS [64] at $\sqrt{s} = 13.6\,$TeV energy in the centre-of-mass

frame. The effects of the $t\bar{t}$-tW interference in the inclusive measurement can be seen to have a small/moderate impact in the signal strength.

The result is higher ($\approx 10\%$) than the already mentioned prediction at aNNLO in QCD ($\sigma_{\text{tW-ch.}} = 71.7 \pm 1.8\,(\text{scale}) \pm 3.4\,(\text{PDF},\alpha_S)$ pb $= 71.7 \pm 3.8$ pb), used for the normalisation of the signal process in the analysis. However, the observation agrees with the expectation within uncertainties. The agreement is superior with the latest prediction, at aN^{3}LO in QCD ($\sigma_{\text{tW-ch.}} = 79.3^{+1.9}_{-1.8}\,(\text{scale}) \pm 2.2\,(\text{PDF},\alpha_S)$ pb $= 79.3^{+2.9}_{-2.8}$ pb), where both the expectation and the observation lie, quite close, within the uncertainties of each other.

Agreement can be found with the latest measurement at the same energy and channel from ATLAS [195]. Recently, the first observation has been done in the semileptonic decay channel by CMS [196], whose result agrees within uncertainties with the presented in this document. These two other analyses also find a larger amount of signal than the prediction at aN^{3}LO and are systematically dominated (though with a larger statistical uncertainty, in the case of the former). Among the uncertainties they consider, in both the JES correction sources play a relevant role in the precision.

The previous inclusive measurement of the same channel and energy in the centre-of-mass frame done by CMS, but with data only from 2016, yielded $63.1 \pm 1.8\,(\text{stat.}) \pm 6.4\,(\text{syst.}) \pm 2.1\,(\text{lumi.})$ pb [3]. This value, consistent with the SM prediction at aNNLO in QCD, lies slightly over one sigma difference with the observation depicted in this document, and over two with the latest expectation at aN^{3}LO, also in QCD. The precision, nevertheless, is relatively similar, with a total $\approx 11\%$ compared with the $\approx 10\%$ obtained in this thesis. Given the common ground between both analysis, and that both are systematically dominated, the modest precision improvement and the difference in observation can draw attention, as apparently, the only remarkable difference should be the increase in luminosity. However, there are various developments between the 2016-only analysis and the full Run 2 measurement that can play a role in the explanation of these differences:

Amount of data used As already mentioned, there is a factor ≈ 4 of difference in the integrated luminosity between the both measurements. This should not affect significantly the final result (although it indeed reduces the statistical uncertainty from a relative 2.9% to 1.1%), as the analyses are systematically limited. However, a larger data set affects indirectly the measurement, as the measurements of corrections (SFs, efficiencies, JES, JER...) can profit from it by reducing their statistical uncertainty.

Updates in detector Beginning from 2017, CMS was upgraded with a new barrel pixel layer and additional forward disk detectors for its tracker subsystem [197]. This improved the reconstruction process of CMS, affecting to precision in the reconstructed observables, such as the impact parameter (thus having effect on b-tagging).

Updates in corrections One significant difference is the change of b-tagging algorithm: from
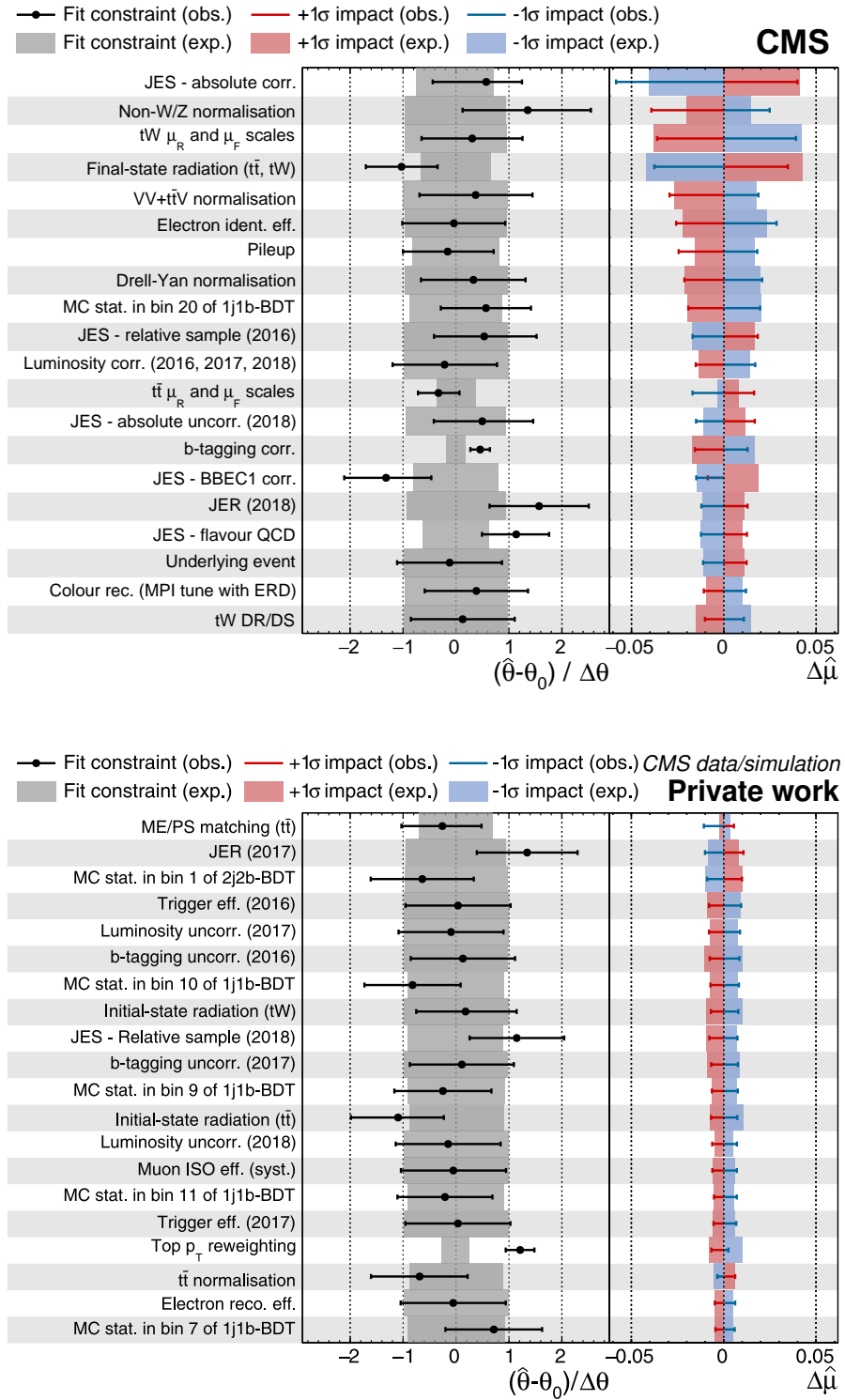
Figure 6.7: The 40 (first half top, second one bottom) largest impacts (right columns) and pulls (middle columns) of the nuisance parameters listed in the left columns from the fit used to determine the inclusive tW cross section. The horizontal bars on the pulls show the ratio of the uncertainties of the fit result to the previous ones, effectively giving the constraint on the nuisance parameter. The label "corr." refers to the correlated component of the uncertainty over the three years and "uncorr." the uncorrelated component for each year. The entries are defined and detailed in Sect. 5.

CSVv2 [178] to `DeepJet`, with better overall performance, as well as the derivation of new corrections for it for the three years. The energy corrections have also been updated in the full Run 2 analysis: a comparison for simulated reconstructed jets of $p_\mathrm{T} = 30$ GeV is shown in Fig. 6.8, where the total JES correction uncertainty is shown as a dependence of η.
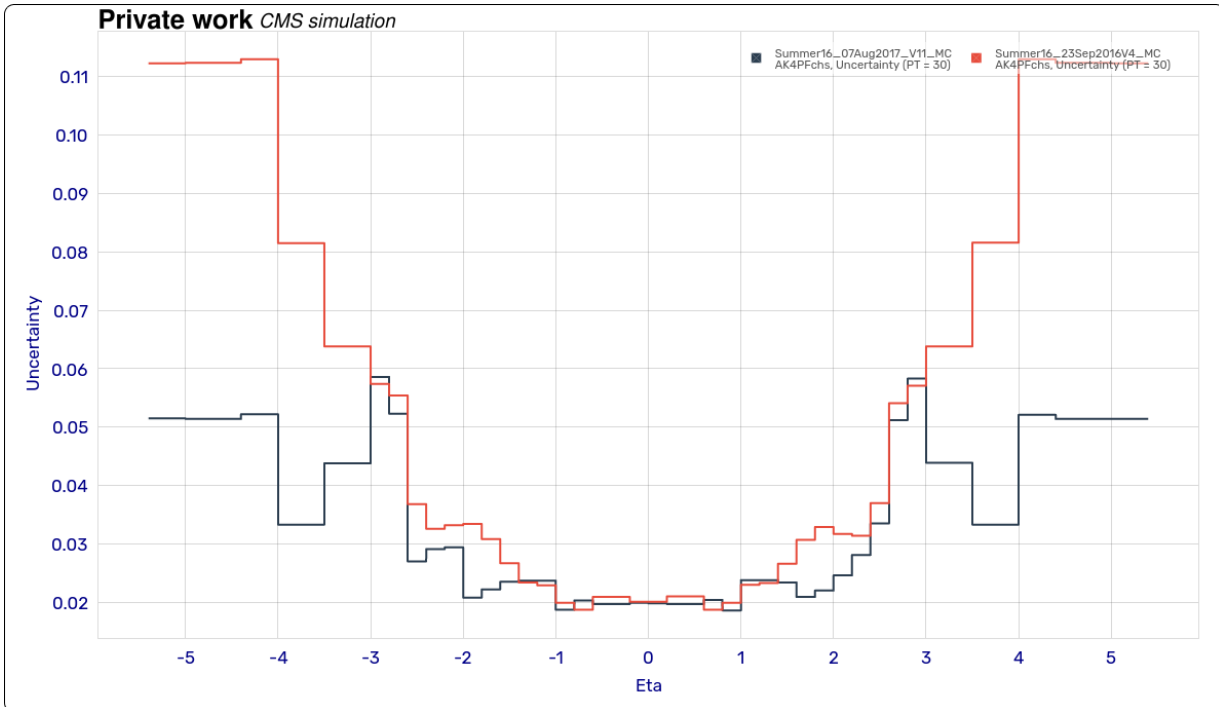
Figure 6.8: Comparison of total JES correction uncertainties for simulated reconstructed jets of $p_\mathrm{T} = 30$ GeV between those used in 2016-only tW inclusive measurement [3] (orange) and the corrections applied for the analysis of this thesis (black).

Updates in luminosity measurement The luminosity uncertainty used in the 2016-only analysis was the best available at the moment: an overall 2.5% precision [198]. However, new measurements from the Collaboration allow to improve the total uncertainty of the 2016-2018 period to 1.6% as mentioned before, and for the 2016 year to 1.2%. This explains the reduction in the uncertainty on the tW cross section: from a 3.3% to 1.5%. The new measurement also slightly varies the luminosity value for 2016 (from 35.9 fb^{-1} to 36.3 fb^{-1}).

Updates in simulation modelling The MC samples in the 2016-only analysis used the CUETP8M1 tune, while this option is only used in the full Run 2 measurement in some background processes of the 2016 year. In addition, the PDF set is different for most of the processes in this thesis observation, using the NNPDF3.1 set, while all simulations in the previous measurement use the NNPDF3.0 [199]. Both differences affect also the uncertainties.

Updates in uncertainty estimation

- The correlation of effects in some sources is different between the 2016-only and full Run 2 measurements, in coherence with an improvement in methodological procedures and recommendations in the CMS Collaboration. An example of such is the leading impact in both analyses: the JES correction source. In the case of the 2016-only measurement, there is only one nuisance present, while in the observation described in this document, the leading impact is actually one nuisance of a set of 21. Other examples are the electron corrections (correlated in a single nuisance in 2016 measurement, separated in ID and reconstruction correction effects in full Run 2), the muon corrections (similar to electrons), or ISR (correlated between the tW and $t\bar{t}$ processes in the previous measurement, separated among those events in this thesis because of the modelling dependance on the process for those emissions).

- The FSR uncertainty is estimated differently. Whereas the current analysis takes the variation obtained with a factor 2 in the energy scale used in the PS to simulate those emissions, the previous one divided that separation by $\sqrt{2}$ and used that as its estimation.

- The ME/PS matching uncertainty was estimated for both $t\bar{t}$ and tW in the 2016-only measurement, while recent CMS observations (such as this thesis' analysis) prefer to only consider the effect on $t\bar{t}$ only while a more profound assessment on the matching uncertainty on tW is done.

- The colour reconnection effects are considered in the full Run 2 by taking advantage of developments done since 2016 in the modelling of this feature. Thus, the connection models considered are different, as the ways of adding them to the maximum likelihood fit: the largest variation with respect to the nominal expectations is taken, per bin, in the 2016-only measurement, whereas this thesis' measurement adds each model as a shape nuisance to exploit the correlations among them.

- Developments in the PS generator (`Pythia 8`) proved that the estimations of the ISR uncertainties are different between the simulations used in the 2016-only analysis and those used by the full Run 2 measurement, being the latter smaller than the former.

- The b-fragmentation and semileptonic B meson decay fraction uncertainty sources are not considered in this analysis. Studies done in CMS showed that the estimation of the b-fragmentation uncertainty source used in the 2016-only analysis could lead to unphysical large values, and are thus not employed in the latest measurements from the Collaboration.

- As mentioned before, the uncertainty in luminosity, PDFs, and of the UE is different among analyses.

Updates in methodology

- Thanks to the increase in data, a larger number of bins in the distributions used for the signal extraction in the three regions is used in the full Run 2 measurement with respect to the 2016-only analysis.

- The training of the MVAs used was optimised to take into account the updated modelling in the input variables as a consequence of the differences with respect to 2016-only analysis (such as the previous ones).

- An extensive and manual source per source, variation per variation, and process per process, review (and treatment, if needed) of all uncertainty sources variations was done in the analysis depicted in this document, as detailed in Sect. 6.2. This was done to increase the fit's stability, and to avoid bias in the fit due to small fluctuations of any template estimation.

The inclusive measurement presented in this chapter thus profits from the advance in the developments in MC modelling, the detector's performance, and of the methodologies used. This provides an updated result, that has the best precision in this process up to date.
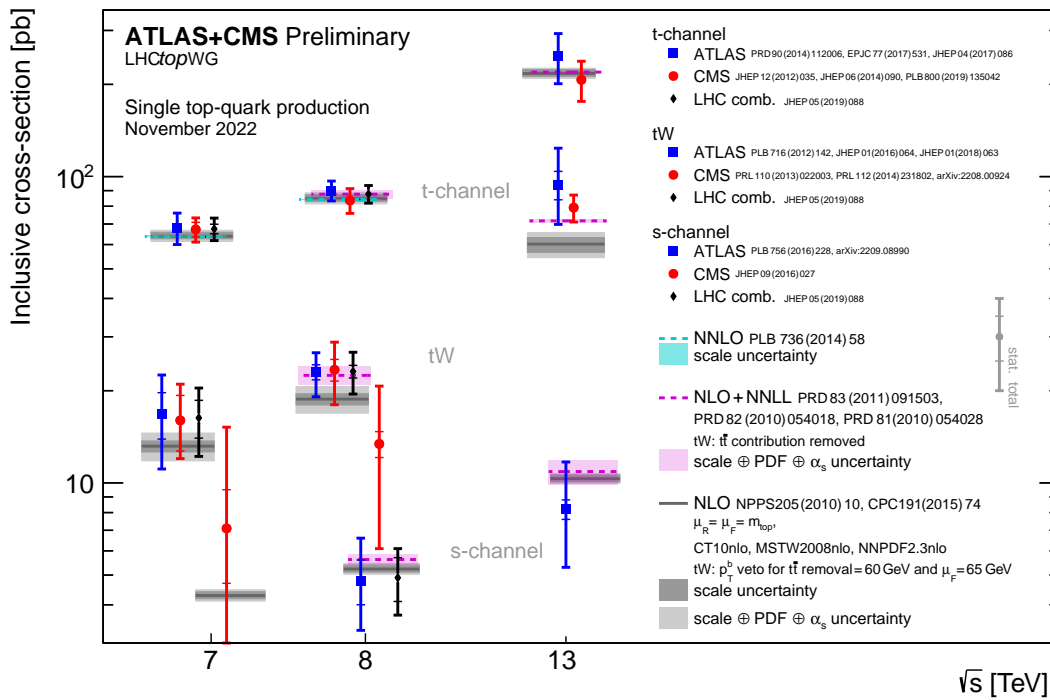
Figure 6.9: Summary of the single top production processes cross sections measured at a centre-of-mass energy of $\sqrt{s} = 13$ TeV with proton-proton collisions.

A summary of all tW measurements at a centre-of-mass energy of $\sqrt{s} = 13$ TeV with proton-proton collisions can be seen in Fig. 6.10, and a larger comparison with the other single top

process in Fig. 6.9.
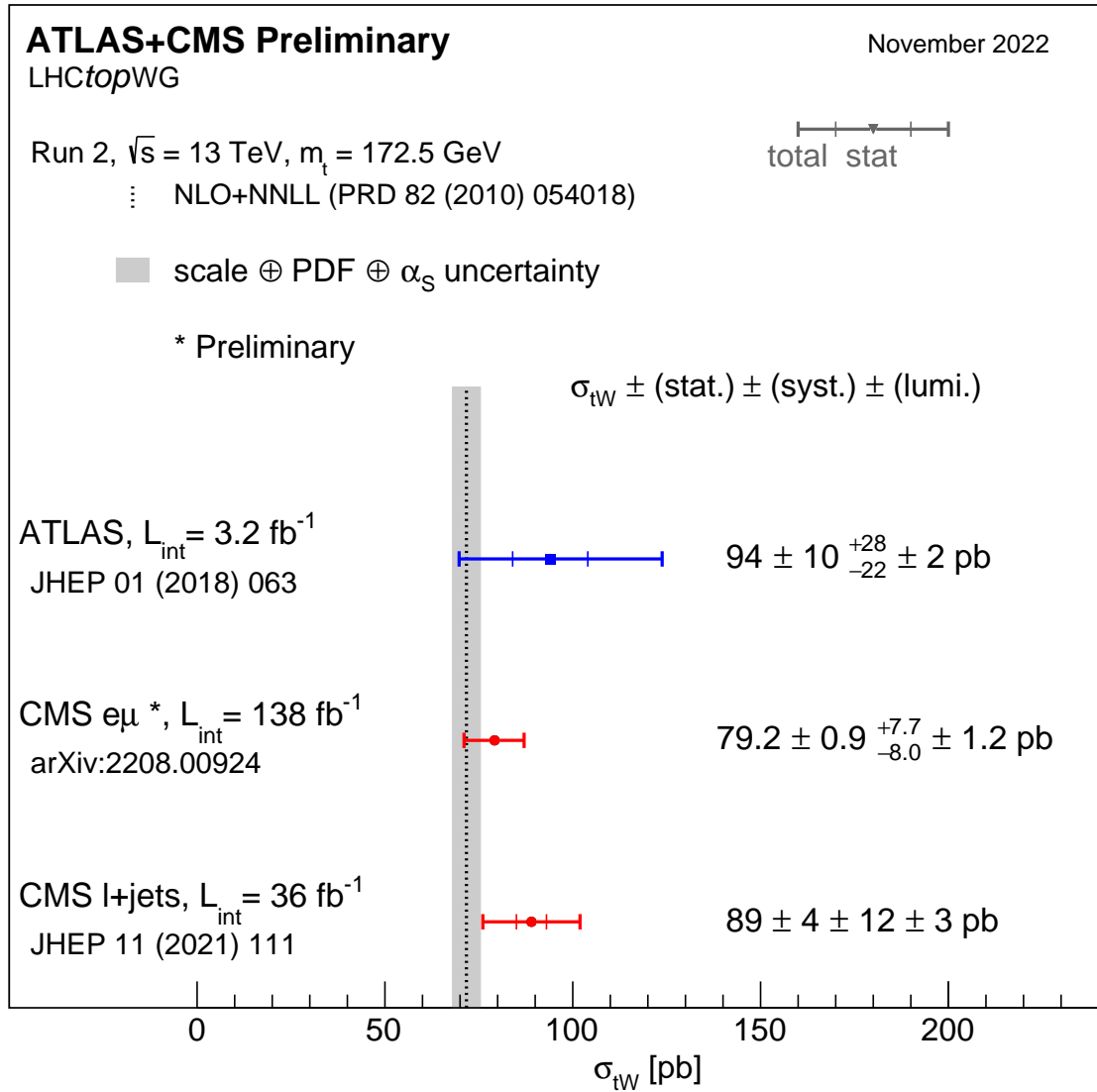
Figure 6.10: Inclusive cross section measurements of the tW process at a centre-of-mass energy of $\sqrt{s} = 13$ TeV with proton-proton collisions in both dileptonic as well as semileptonic channels.

7 Differential measurement

M EASUREMENTS of the cross section as a function of observables (or differential measurements) can be seen as a second step after studying the inclusive value. Such results need more data to probe the cross section in bins of different observables. In return, one can analyse the modelling of variables by different generators, and search for deviations from the SM that affect the shape of observables, instead of only its normalisation. To accomplish differential measurements, one needs to consider the potential variations on the chosen observable due to the detector effects. These alterations are removed from data through a procedure called unfolding, that is described now. Afterwards, the differential result is considered: first the event selection and then the values obtained.

7.1 Unfolding

The variables of the particles we store, such as the transverse momentum of one muon, do not reflect exactly the real features of the particles we measure. The reason is that our detector affect the measure itself, as its components interact with the particles, and consequently our measurements are smeared (e.g. measuring a muon of $20\,\mathrm{GeV}$ as one of $35\,\mathrm{GeV}$). Other possibility is that our detector, although the particle do passes through it, does not detect it. And on top of these effects, one must take into account that each measure has its own uncertainty and that as we are seeing these distributions of the variables from histograms, we depend also in the amount of data that we have.

When one tries to measure a total cross section, the effects of changing the value of the variable are not relevant, because the crucial information is the total amount of signal that we have. The effects of not reconstruct particles are taken into account because they affect the total amount of signal. Otherwise, in differential measurements, both effects are important, because it becomes necessary to know the amount of events that one have with (e.g.) an electron transverse momentum of $20 - 30\,\mathrm{GeV}$ and not of $30 - 50\,\mathrm{GeV}$. The procedure of removing those effects is called unfolding.

The problem can be easily presented as follows. Let us set the physical values of some variable in some bin i as μ_i and the measured and stored values as ν_i. One can parametrise the effects of the detector in a matrix R called the **response matrix**. Thus:

$$\vec{\nu} = R\vec{\mu}, \quad \nu_i = R_{ij}\mu_j \tag{7.1}$$

These matrices are a feature of our detector and our methods of reconstructing events, that here are crucial. They can be obtained through simulations that only contain signal events whose characteristics are those that we expect that our *actual* data have. To estimate them, one must fix the generator information level up to which the unfolding is to be done. In this thesis, we will unfold data to particle level from detector (or reconstruction) level. Then, we can estimate the response matrix simply as

$$R_{ij} = \frac{n_{ij}}{n_i}, \tag{7.2}$$

where n_{ij} are the number of reconstructed events whose value of the measured variable fall in the detector-level bin j that had its particle-level value in the bin i, and n_i are the number of events whose particle-level value of the variable fell in the bin i. The number of bins (and their limits) in the detector level does not have to be the same as in particle level.

Starting from the problem defined in Eq. (7.2), it is trivial to guess a very direct solution: invert the response matrix so that one can obtain the values at particle level. Unfortunately, inverting matrices can be a very challenging numerical problem depending, of course, on the matrix itself: the more diagonal the matrix is, more easy numerically is to be inverted. There is a way to enhance the "diagonality" of R: the choice of binning at detector and particle level. From the previous definitions, two quantities can be defined that will be useful later: the **stability** of a particle-level bin i and the **purity** of a detector-level bin j. They are defined respectively as follows,

$$s_i := \frac{\sum\limits_{j=1}^{N_{\text{bins}}^{\text{detector}}} n_{ij}}{n_i} \qquad p_j := \frac{\sum\limits_{i=1}^{N_{\text{bins}}^{\text{particle}}} n_{ij}}{n_j^R}, \tag{7.3}$$

where n_{ij} and n_i are the same mentioned before and where n_j^R is the amount of simulated events in the (detector-level) bin j. Essentially, stabilities give us a notion of, on one hand, the amount of simulated events that we end up reconstructing and selecting as signal, but also of how many of them stay in the same bin and are not measured elsewhere. Purities are an estimation of the amount of reconstructed events in one bin j related with the signal process over the total number of reconstructed events in that bin. The relevant point is that the maximisation of them through the choice of binning in the detector and particle levels enhances the diagonality of the response matrix, thus making the unfolding problem easier.

The common procedure to perform the unfolding, however, is not that of directly inverting the matrix. It can be shown that the problem can be rewritten as finding the values $\vec{\mu}$ that minimise a χ^2 expression such as

$$\chi_R^2(\vec{\mu}) = (R\vec{\mu} - \vec{\nu})^T V^{-1}(R\vec{\mu} - \vec{\nu}). \tag{7.4}$$

In this expression, V^{-1} represents the covariance matrix that ultimately encodes the uncertainties of our final measurements. At the end, we face again other minimisation problem. This allows to implement a way of dealing with problems that might arise when the response matrix is not very diagonal (a feature that, as already said, is undesirable) that is called **regularisation**. In practice, this characteristic implies to add a new term to the previous χ^2 that modifies effectively the minimisation: the difference between the various regularisation approaches is how this new term is defined. If one were to do such thing, the first conclusion would be that the final result is affected, as all regularisation adds an artificial bias in the whole procedure. However, such a thing might be preferred when the alternative is the impossibility of unfolding at all.

In general, all regularisations are modulated by a parameter τ that allows us to get a final χ^2 expression for the minimisation as

$$\chi_{\text{unf.}}^2(\vec{\mu}, \lambda) = \chi_R^2(\vec{\mu}) + \tau \cdot \chi_{\text{reg.}}^2(\vec{\mu}) + \lambda \sum_i (R\vec{\mu} - \vec{\nu})_i \tag{7.5}$$

Here, we have added another term (the last) that is necessary to account for some problems that can arise when bins with a low amount of events are present. If that were to happen, the count of those events would follow a Poisson distribution effectively, not a Gaussian, and for the χ^2 minimisation approach it is necessary that a Gaussian distribution is followed in all bins. To correct for those possible divergences, that last term is added, which helps to take into account the total amount of events (the normalisation). This is term is denoted as an **area constraint**. The choice of the parameter τ allows to tune the regularisation effect. It is commonly chosen as the best compromise between bias and variance in the final result through the L-curve method [200].

This known approach, used in various differential analyses, is implemented through e.g. `TUnfold` [201], and has been used in various Run 2 analyses (e.g. [2, 30]). However, another alternative method to unfold the data can be done using a maximum likelihood fit, and merging into one step the signal extraction and the unfolding. In this approach, a fit like the one from Sect. 6.2 is used, but with a modified likelihood: instead of having just the signal strength as one POI, we have the number of particle-level bins, $N_{\text{bins}}^{\text{particle}}$ of POIs, each of them being the signal strength of the i particle-level bin. The idea behind is to separate the signal process template into multiple ones, having one template per particle-level bin and one extra that contains all MC signal events in the signal region that do not fulfill the requirements for the fiducial region (at particle level). To obtain all of them, we use the signal-process simulations. At the end, we can derive from the signal strength in each particle-level bin the associated contributions, which would be the result.

7.2 Event selection

The signal region used for the differential measurement is defined upon the baseline selection but unlike the inclusive observation, we only use 1j1b events, that show the largest signal contribution of all the bins. In addition, requiring exactly one b-tagged jet reduces the potential to have events from the doubly-resonant diagrams. Those events, more affected by the interference between processes, are expected to have a larger jet multiplicity. The differences between the various models used to treat the interference are expected to be higher when the presence of events from doubly-resonant diagrams is larger, and vice versa. The selection is finally further purified (in terms of signal presence) by vetoing the presence of loose jets, increasing this way the signal-to-background ratio as seen in Fig. 7.1. The data and event expectations of the signal region, 1j1b $+0j_l$ are shown in Table 7.1.

Process	1j1b $+0j_l$
tW	$17\,600 \pm 1\,500$
$t\bar{t}$	$59\,000 \pm 9\,000$
DY	$2\,400 \pm 300$
$VV + t\bar{t}V$	$1\,100 \pm 500$
Non-W/Z	240 ± 110
Total	$80\,000 \pm 9\,000$
Data	$75\,200 \pm 300$

Table 7.1: The number of observed and MC predicted events in the differential measurement signal region, 1j1b $+0j_l$ region. The statistical uncertainties in the data and the total uncertainties in the predictions are given.

The identification of particle-level objects is summarised in Table 7.2. Objects are defined following Sect. 3.2.1 guidelines. A fiducial region is built as described in Table 7.3, where particle-level jets are also *cleaned* the same way as their detector-level equivalents, but using dressed leptons (as described in Sect. 3.2.1). For the unfolded distributions in the fiducial region, as shown later in the results of Figs. 7.4 and 7.5, all MC simulations show very similar shapes. Therefore, this choice of fiducial region reduces these effects and the corresponding modelling uncertainty associated with the interference treatment (see Section 5).

Object	p_T (GeV)	$|\eta|$
Muons	>20	<2.4
Electrons	>20	<2.4, excluding [1.444–1.566]
Jets	>30	<2.4
Loose jets	$>20, <30$	<2.4

Table 7.2: Selection requirements for particle-level objects.
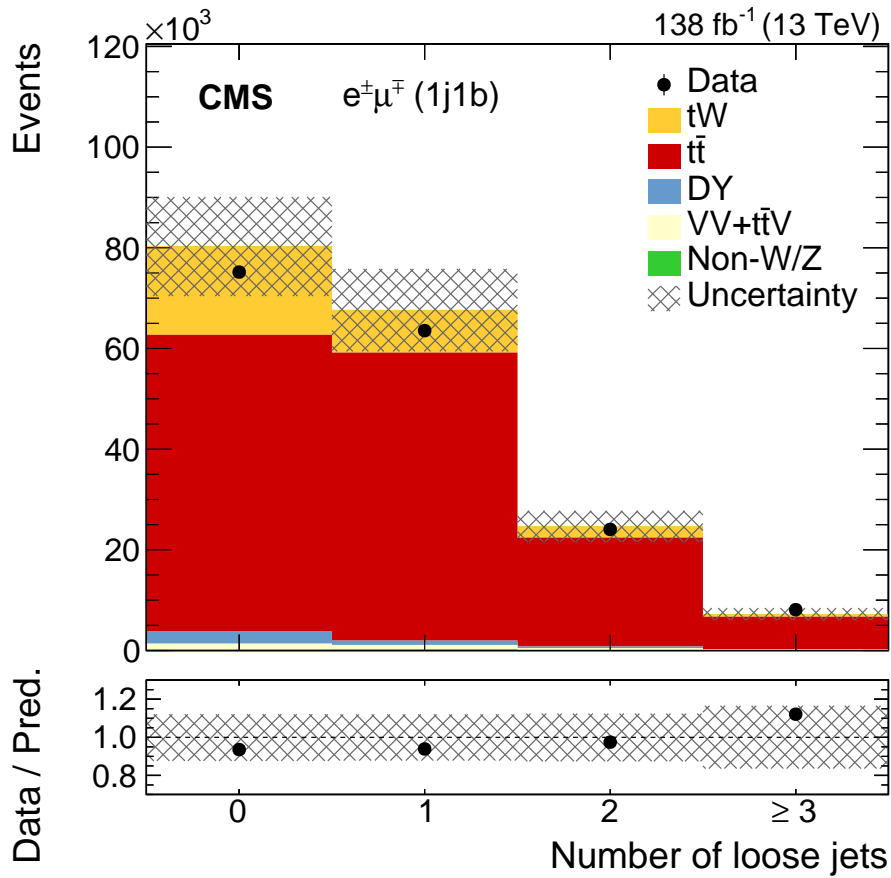
Figure 7.1: Data/MC comparison of the number of loose jets of 1j1b events. Vertical bars on the points show the statistical uncertainty in the data, and the hatched band the total uncertainty from the fitted expectations. The lower panels show the ratio of the data to the sum of post-fit MC.

Variable	Criterium
Number of leptons	≥ 2
Leading lepton p_T	>25 GeV
Invariant mass of all dilepton pairs	>20 GeV
Number of jets	1
Number of loose jets	0
Number of b jets	1

Table 7.3: Definition of the fiducial region.

7.3 Signal extraction

The signal extraction and the unfolding for the fiducial differential cross section measurement are performed with an maximum likelihood fit. This choice of approach was done during the internal review of the analysis in the Collaboration, by means of a thorough study of various alternatives. These were variations between a background subtraction and unfolding performed with the `TUnfold` package mentioned in Sect. 7.1 to a joint signal extraction plus unfolding in a maximum likelihood fit, with intermediate options such as maximum likelihood fits for the signal extraction with `TUnfold` unfolding (propagating uncertainties with the fit's covariance matrix). The internal review concluded that the best option was a combined maximum likelihood fit to do both, based on better expected (using the Asimov dataset) results.

   The fit, whose technical details are the same as those of Sect. 6.2, has some differences to accommodate the differential measurement's needs. The POIs are the strengths of the signal process in each bin of the particle-level distribution. The signal sample is divided into as many contributions as there are particle-level bins. There is a 7% contribution to the signal region from non-fiducial events. We treat these events as a background so the strength associated with them is not a POI. One nuisance parameter for each systematic uncertainty source is added to the fit.

   The differential cross section is measured as a function of the following physical observables:

- leading lepton p_{T};

- jet p_{T};

- $\Delta\varphi(e^{\pm}, \mu^{\mp})$: the azimuthal angle difference between the two leptons;

- $p_z(e^{\pm}, \mu^{\mp}, \mathrm{j})$: the longitudinal momentum component of the dilepton + jet system;

- $m(e^{\pm}, \mu^{\mp}, \mathrm{j})$: the invariant mass of the dilepton + jet system; and,

- $m_{\mathrm{T}}(e^{\pm}, \mu^{\mp}, \mathrm{j}, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$: the transverse mass of the dilepton + jet + $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ system. For a collection of particles with transverse momentum $\vec{p}_{\mathrm{T},i}$, m_{T} is defined as:

$$m_{\mathrm{T}} = \sqrt{\left(\sum_i |\vec{p}_{\mathrm{T},i}|\right)^2 - \left|\sum_i \vec{p}_{\mathrm{T},i}\right|^2}. \tag{7.6}$$

The first two variables shown above provide information on the kinematic properties of the events. The $\Delta\varphi(e^{\pm}, \mu^{\mp})$ variable probes the kinematic and polarisation correlations between the top quark and W boson. The p_z distribution can be used to study the boost of the tW system. The last two variables, the dilepton + jet invariant mass and m_{T}, are sensitive to the invariant mass of the tW system. The distributions from the data and simulation for these six variables in the differential signal region are shown in Fig. 7.2. As in the case of the distributions for the inclusive measurement, overall there is good agreement within the uncertainties between the data and simulation, though the data are consistently lower than the predicted values. The binning and

Variable	Leading lepton p_T	Jet p_T	$\Delta\varphi(e^\pm,\mu^\mp)/\pi$	$p_z(e^\pm,\mu^\mp,\mathrm{j})$	$m_\mathrm{T}(e^\pm,\mu^\mp,\mathrm{j},\vec{p}_\mathrm{T}^{\,\mathrm{miss}})$	$m(e^\pm,\mu^\mp,\mathrm{j})$
Condition number	2.92	4.69	1.03	2.06	5.27	2.82

Table 7.4: Condition numbers from the response matrices of the variables unfolded in the differential measurement.

number of the variables in the fit was selected by aiming to have constant and large purities and stabilities, as well as sufficient expected events.

The measurement is performed using the three years of data taking. The detector response, estimated with these matrices, is similar for the three years, after comparing the condition number from all as well as ratios across years. Thus, the measurement is performed directly using the combined data set, and its associated response matrices can be seen in Fig. 7.3. Although the method used for unfolding is a maximum likelihood fit, where the response matrices are unrolled in the distribution through the use of various POIs, they are still useful to estimate the migrations across bins due to the detector effects. Thus they can be used to foresee potential issues in the fit, due to the fact of having several particle-level signal contributions in one unique bin at detector level. Their condition numbers are shown in Tab. 7.4, and given they round $O(10^0)$ values, we can expect few complications in the fit because of this and unfolding without regularisation is most probably feasible.

The same checks, automatic and not, done for the inclusive measurement fit and detailed in Sect. 6.2 are done for the differential ones. In this case, an extra complication is found, as we have more templates to extract: all the templates extracted for the signal in the inclusive process are now derived for each one of the particle-level bins of each variable, in addition to the estimation from nonfiducial events. On top of these checks, others were done with the Asimov data set to ensure the closure of the method (by comparing with the particle-level generated information after unfolding, or using randomly-generated data from it).

After the maximum likelihood fit, the result is normalised to the fiducial cross section (obtained from the summation of the contents of the bins), and the bin width. The uncertainties are propagated taking into account the correlations across bins after the fit and with the estimation of the fiducial cross section itself. It is easy to prove that

$$Var(d\sigma_{fid.}(i)) = \frac{Var(d\sigma(i))}{\sigma_{fid.}^2} + \frac{Var(\sigma_{fid.})\cdot(d\sigma(i))^2}{\sigma_{fid.}^4} - \frac{2(d\sigma(i))}{\sigma_{fid.}^3}\cdot Cov(d\sigma(i),\sigma_{fid.}), \qquad (7.7)$$

where i represents one bin of the differential cross section, $Var(x)$ the variance of the random variable x, $d\sigma_{fid.}(i)$ the differential cross section normalised to the fiducial cross section in the bin i, $d\sigma(i)$ the differential cross section in the bin i and $\sigma_{fid.}$ the fiducial cross section.

Similarly, to obtain the complete covariance matrix from the normalised result we use the following expression (derived in the same way), which trivially is a more general case than the
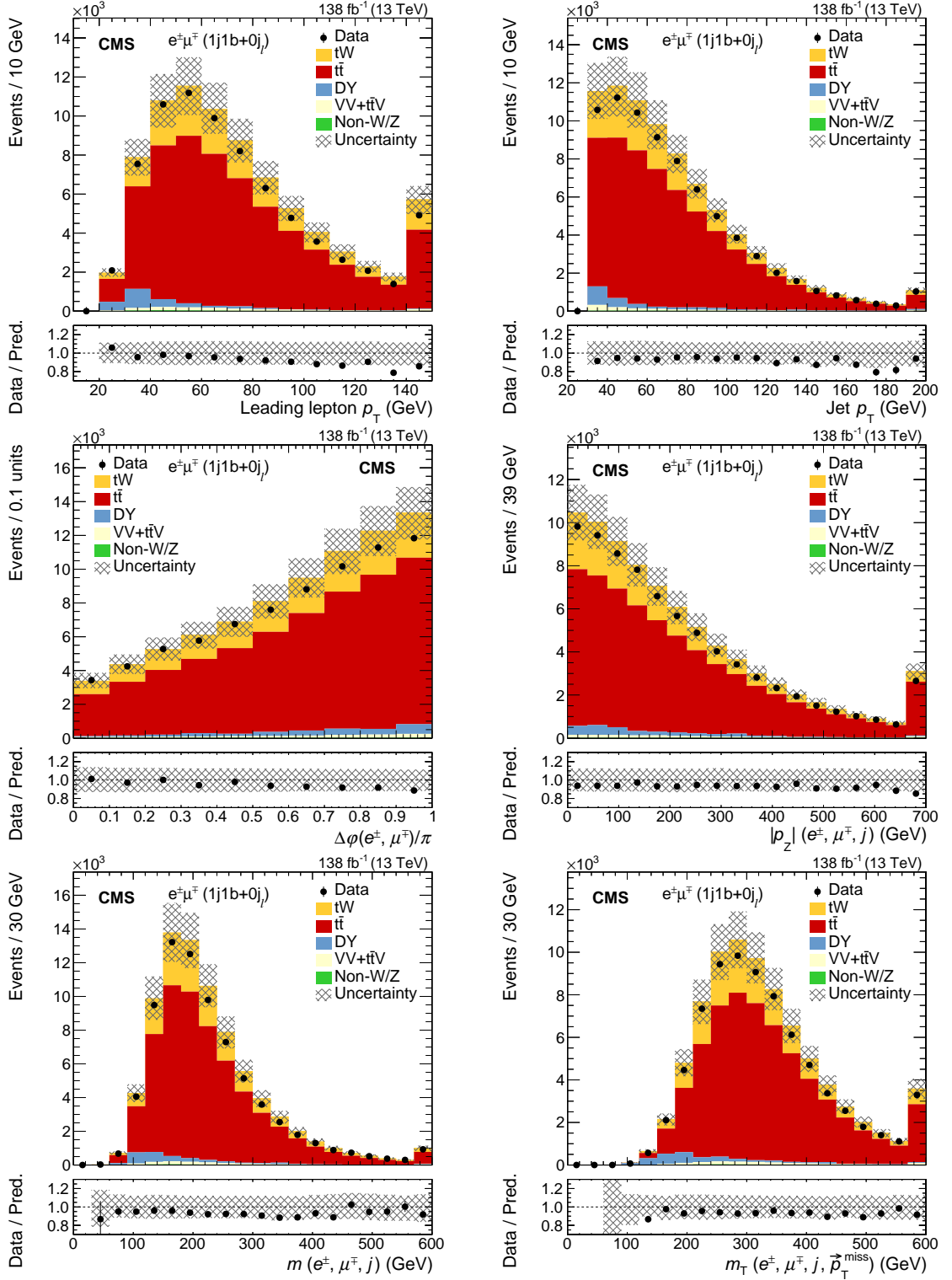
Figure 7.2: The measured distributions from data (points) and MC simulations (coloured histograms) of the six observables used to measure the tW differential cross sections. The last bin of each distribution contains the overflow events. The vertical bars on the data show the statistical uncertainty. The hatched band displays the sum of the statistical and systematic uncertainties in the MC predictions before the fit. The lower panels show the ratio of the data to the sum of the MC expectations. The MC simulations are normalised to their theoretical cross section values as described in Section 3.3.
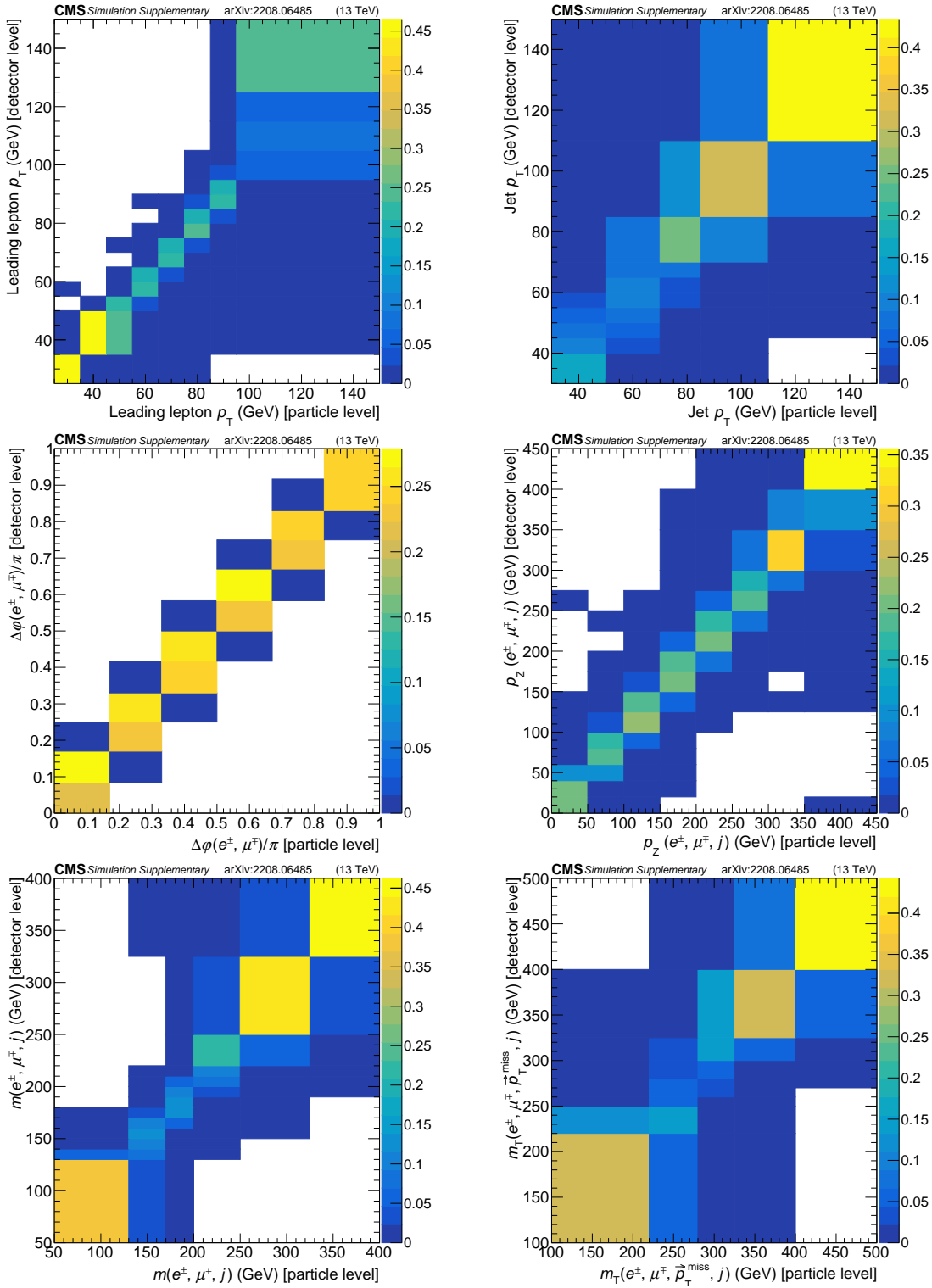
Figure 7.3: Response matrices extracted from MC simulations of the six observables used to measure the tW differential cross sections. Signal events in the 1j1b +0j$_l$ are selected.

previous one:

$$
\begin{aligned}
Cov(d\sigma_{fid.}(i), d\sigma_{fid.}(j)) = {} & \frac{Cov(d\sigma(i), d\sigma(j))}{\sigma^2_{fid.}} + \frac{Var(\sigma_{fid.}) \cdot d\sigma(i) \cdot d\sigma(j)}{\sigma^4_{fid.}} \\
& - \frac{(d\sigma(i))}{\sigma^3_{fid.}} \cdot Cov(d\sigma(j), \sigma_{fid.}) - \frac{(d\sigma(j))}{\sigma^3_{fid.}} \cdot Cov(d\sigma(i), \sigma_{fid.}).
\end{aligned}
\tag{7.8}
$$

A natural consequence of this is also that the covariance matrix of the normalised result has one degree of freedom less than its absolute counterpart. This can be seen by exploiting the fact that the absolute result in any i bin $d\sigma(i)$ can always be written as

$$
d\sigma(i) = d\sigma_{fid.} - d\sigma(1) - d\sigma(2) - \ldots - d\sigma(i-1) - d\sigma(i+1) - \ldots - d\sigma(n-1) - d\sigma(n), \tag{7.9}
$$

with n being the number of particle-level bins of the observable.

7.4 Result

The tW differential cross sections, normalised to the total fiducial cross section $\sigma_{\text{fid.}}$ (obtaining by summing the contents of the particle level bins), are shown in Figs. 7.4 and 7.5 from the data and the MC predictions. Figures 7.7, 7.8, 7.9, 7.10, 7.11, and 7.12 show the 20 largest impacts, as defined in the inclusive measurement, after the maximum likelihood fit done to obtain the absolute values for each observable. The uncertainties on the normalised result are roughly 10–50% in most cases. As in the case of the inclusive measurement, they are dominated by the systematic uncertainties, as shown in the majority of the right column plots of Figs. 7.4 and 7.5. In addition, modelling sources are overall predominant, closely followed by experimental ones, while background normalisation and the limited size of MC samples have little effect globally.

Variable	PH DR + P8	PH DS + P8	PH DR + H7
Leading lepton p_{T}	0.02	0.01	0.03
Jet p_{T}	0.14	0.27	0.01
$\Delta\varphi(e^\pm, \mu^\mp)/\pi$	0.26	0.29	0.32
$p_z(e^\pm, \mu^\mp, \text{j})$	0.70	0.77	0.82
$m_{\text{T}}(e^\pm, \mu^\mp, \text{j}, \vec{p}_{\text{T}}^{\text{miss}})$	0.54	0.60	0.59
$m(e^\pm, \mu^\mp, \text{j})$	0.03	0.02	0.28

Table 7.5: The p-values from the goodness-of-fit tests comparing the six differential cross section measurements with the predictions from POWHEG (PH) + Pythia 8 (P8) DR and DS and Herwig 7 (H7) DR. The complete covariance matrix from the results and the statistical uncertainties in the predictions are taken into account.

Tables 7.5 and 7.6 give the p-values from the χ^2 goodness-of-fit tests done for the six distributions, using the different MC generators and taking into account the full covariance matrix of each result, as well as the statistical uncertainties of the MC predictions. The full covariance matrix is
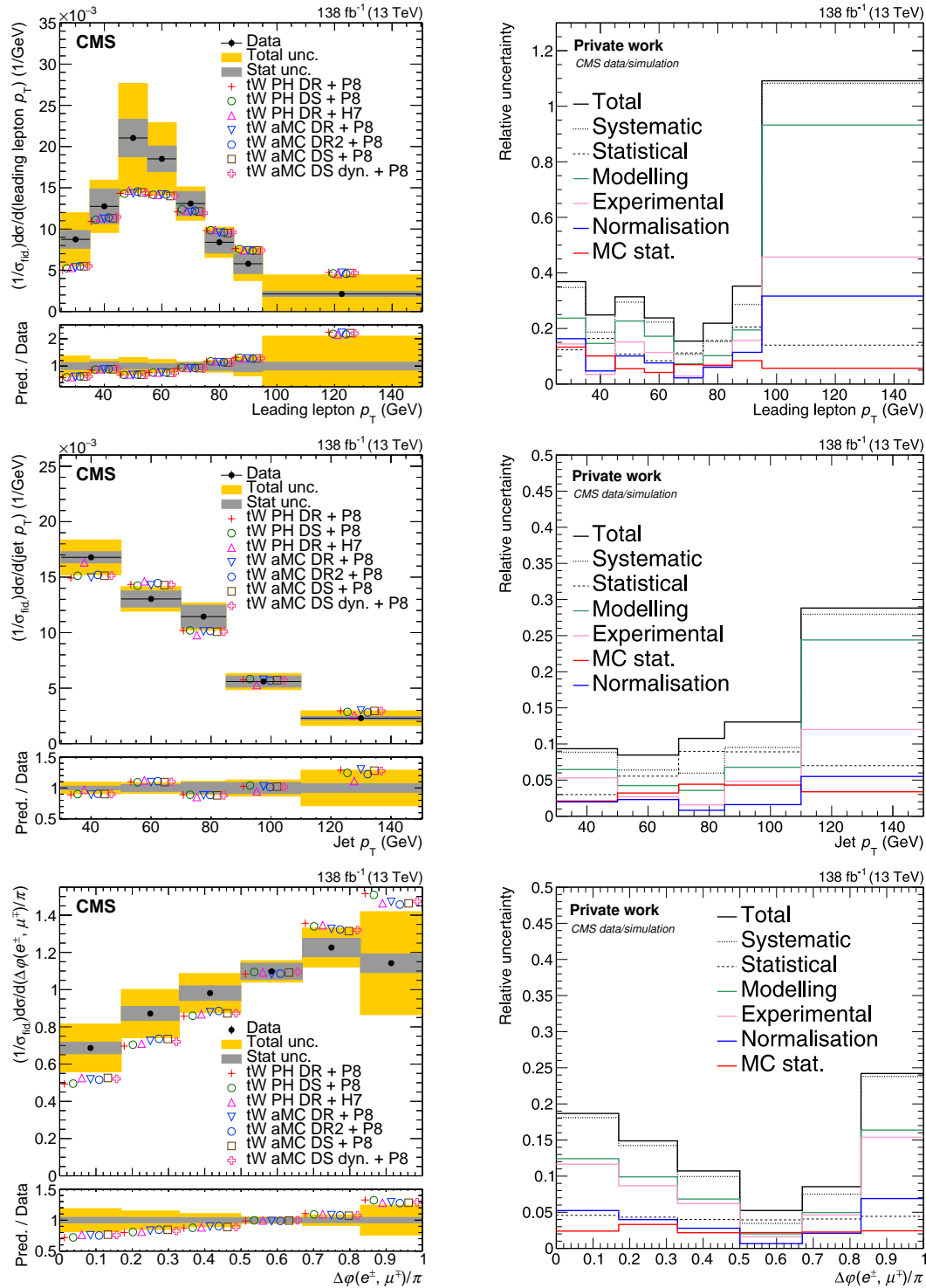
Figure 7.4: Normalised fiducial differential tW production cross section as function of the p_T of the leading lepton (upper left), p_T of the jet (middle left), $\Delta\varphi(e^{\pm}, \mu^{\mp})$ (lower left), and their relative uncertainties (right column). The vertical bars on the points give the statistical uncertainty in the data and the horizontal ones the bin width in the left column plots. Predictions from POWHEG (PH) + Pythia 8 (P8) DR and DS, Herwig 7 (H7) DR, MadGraph5_aMC@NLO (aMC) + Pythia 8 DR, DR2, DS, and DS with a dynamic factor are shown. The grey band represents the statistical uncertainty and the orange band the total one. The lower panels show the ratio of predictions to data. Relative uncertainties are grouped and sorted in the legend by the average of the effect in all bins per set, except for the first three entries.

Figure 7.5: Normalised fiducial differential tW production cross section as functions of the $p_z(e^{\pm}, \mu^{\mp}, j)$ (upper left), $m(e^{\pm}, \mu^{\mp}, j)$ (middle left), $m_T(e^{\pm}, \mu^{\mp}, j, \vec{p}_T^{\,\mathrm{miss}})$ (lower left), and their relative uncertainties (right column). The vertical bars on the points give the statistical uncertainty in the data and the horizontal ones the bin width in the left column plots. Predictions from POWHEG (PH) + Pythia 8 (P8) DR and DS, Herwig 7 (H7) DR, MadGraph5_aMC@NLO (aMC) + Pythia 8 DR, DR2, DS, and DS with a dynamic factor are shown. The grey band represents the statistical uncertainty and the orange band the total one. The lower panels show the ratio of predictions to data. Relative uncertainties are grouped and sorted in the legend by the average of the effect in all bins per set, except for the first three entries.

Variable	aMC DR + P8	aMC DR2 + P8	aMC DS + P8	aMC DS dyn. + P8
Leading lepton p_T	0.05	0.04	0.03	0.07
Jet p_T	0.15	0.11	0.14	0.12
$\Delta\varphi(e^\pm, \mu^\mp)/\pi$	0.33	0.40	0.37	0.32
$p_z(e^\pm, \mu^\mp, \mathrm{j})$	0.76	0.86	0.84	0.82
$m_\mathrm{T}(e^\pm, \mu^\mp, \mathrm{j}, \vec{p}_\mathrm{T}^{\,\mathrm{miss}})$	0.49	0.51	0.48	0.52
$m(e^\pm, \mu^\mp, \mathrm{j})$	0.09	0.12	0.10	0.14

Table 7.6: The *p*-values from the goodness-of-fit tests comparing the six differential cross section measurements with the predictions from `MadGraph5_aMC@NLO` (aMC) + `Pythia 8` DR, DR2, DS, and DS with a dynamic factor. The complete covariance matrix from the results and the statistical uncertainties in the predictions are taken into account.

obtained by normalising the covariance matrix extracted from the maximum likelihood fit to the measured fiducial cross section as described in the previous section: the matrices for the six observables are shown in Fig. 7.6. These tests show a poorer compatibility in the leading lepton p_T, $m(e^\pm, \mu^\mp, \mathrm{j})$, and jet p_T distributions with the nominal `POWHEG` DR prediction than in the other variables. In most of the cases, the *p*-values determined from the distributions of all the variables are similar for the other expectations. When comparing data to the predictions, there is a slight disagreement in the leading lepton p_T and the $\Delta\varphi(e^\pm, \mu^\mp)$ differential cross sections. Other CMS measurements have measured similar tensions in the top quark p_T [194] and $\Delta\varphi(e^\pm, \mu^\mp)$ [202] variables. All methods, DR, DR2, DS, and DS with a dynamic factor, show similar compatibility with the measurements, as well as small differences among them. This is also true for the DR predictions interfaced with `Herwig 7`.

ATLAS' differential measurement of this process at $\sqrt{s} = 13$ TeV [203] obtains relative uncertainties that also vary per bin and observable significantly, as in the results of this chapter. The range of 10–50% total relative uncertainty can be also applied to ATLAS' measurement, as well as the clear systematic domination of the sources. Among them, modelling uncertainties are one of the sources that predominates in roughly all observables and bins, as in our case. There are two distributions that are measured in both analyses: $m(e^\pm, \mu^\mp, \mathrm{j})$ and $m_\mathrm{T}(e^\pm, \mu^\mp, \mathrm{j}, \vec{p}_\mathrm{T}^{\,\mathrm{miss}})$. Differences such as binning or the exact uncertainty sources considered make comparisons hard across measurements. However, the precision achieved in this result is overall slightly better to that of [203], except for the low edges of the $m_\mathrm{T}(e^\pm, \mu^\mp, \mathrm{j}, \vec{p}_\mathrm{T}^{\,\mathrm{miss}})$ variable, where we pay a better resolution (more bins) with higher uncertainties. In both cases, bins span larger in [203] than in this measurement, where we increase our resolution in the bulk of the distributions (low/moderate values). In any case, the most significant divergence between both measurements is the methodology. ATLAS' measurement takes advantage of a MVA classifier to separate tW from $t\bar{t}$ that constrains the set of distributions to unfold without added bias to the observation of the cross section. Our methodology allows us to freely unfold any distribution, such as the leading lepton or jet p_T.
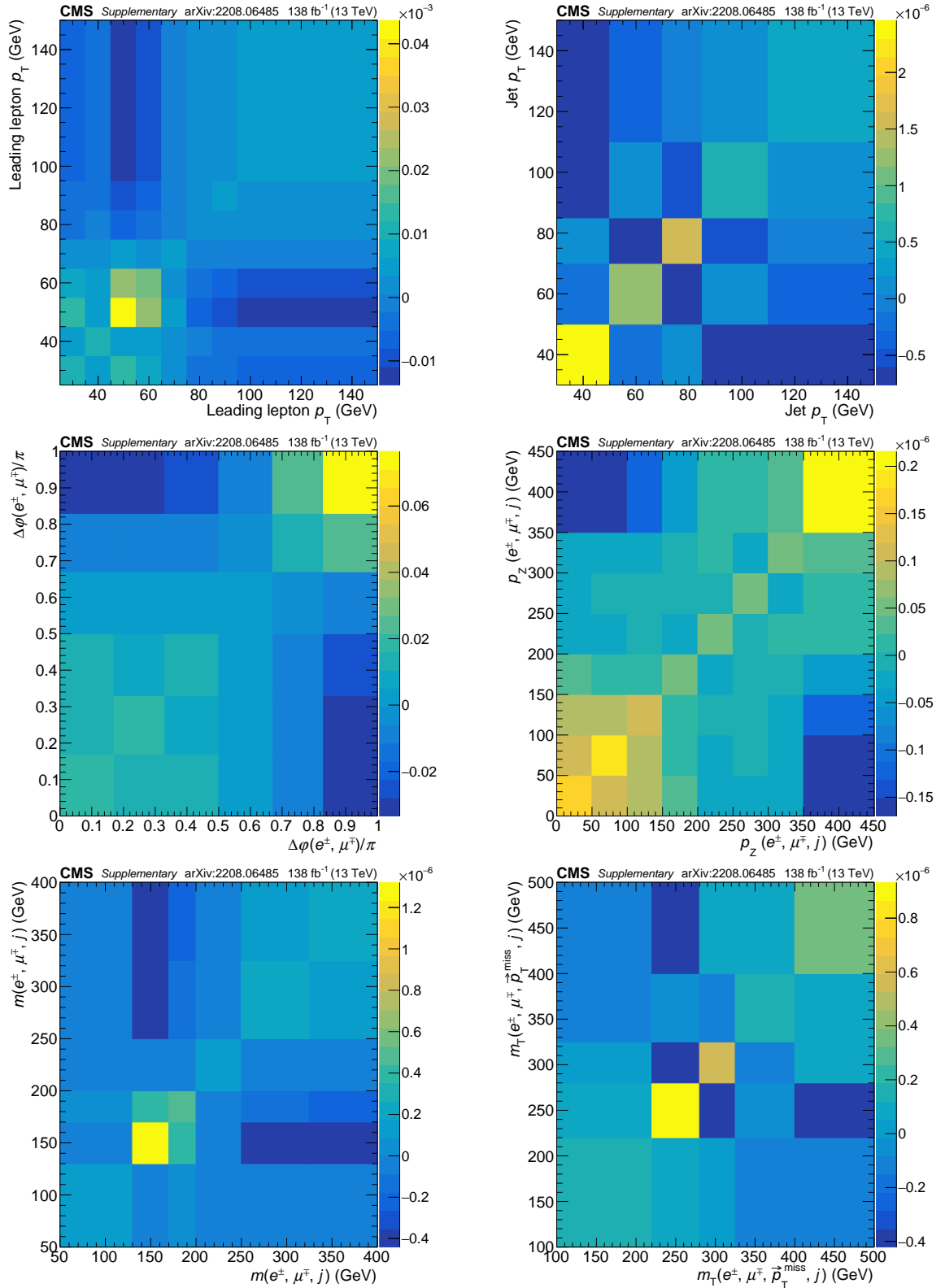
Figure 7.6: Total covariance matrices unfolded to particle-level and normalised to the fiducial cross section (and bin width)as functions of the p_T of the leading lepton (upper left), $p_z(e^\pm, \mu^\mp, j)$ (upper right), p_T of the jet (middle left), $m(e^\pm, \mu^\mp, j)$ (middle right), $\Delta\varphi(e^\pm, \mu^\mp)$ (lower left), and $m_T(e^\pm, \mu^\mp, j, \vec{p}_T^{\text{miss}})$ (lower right).
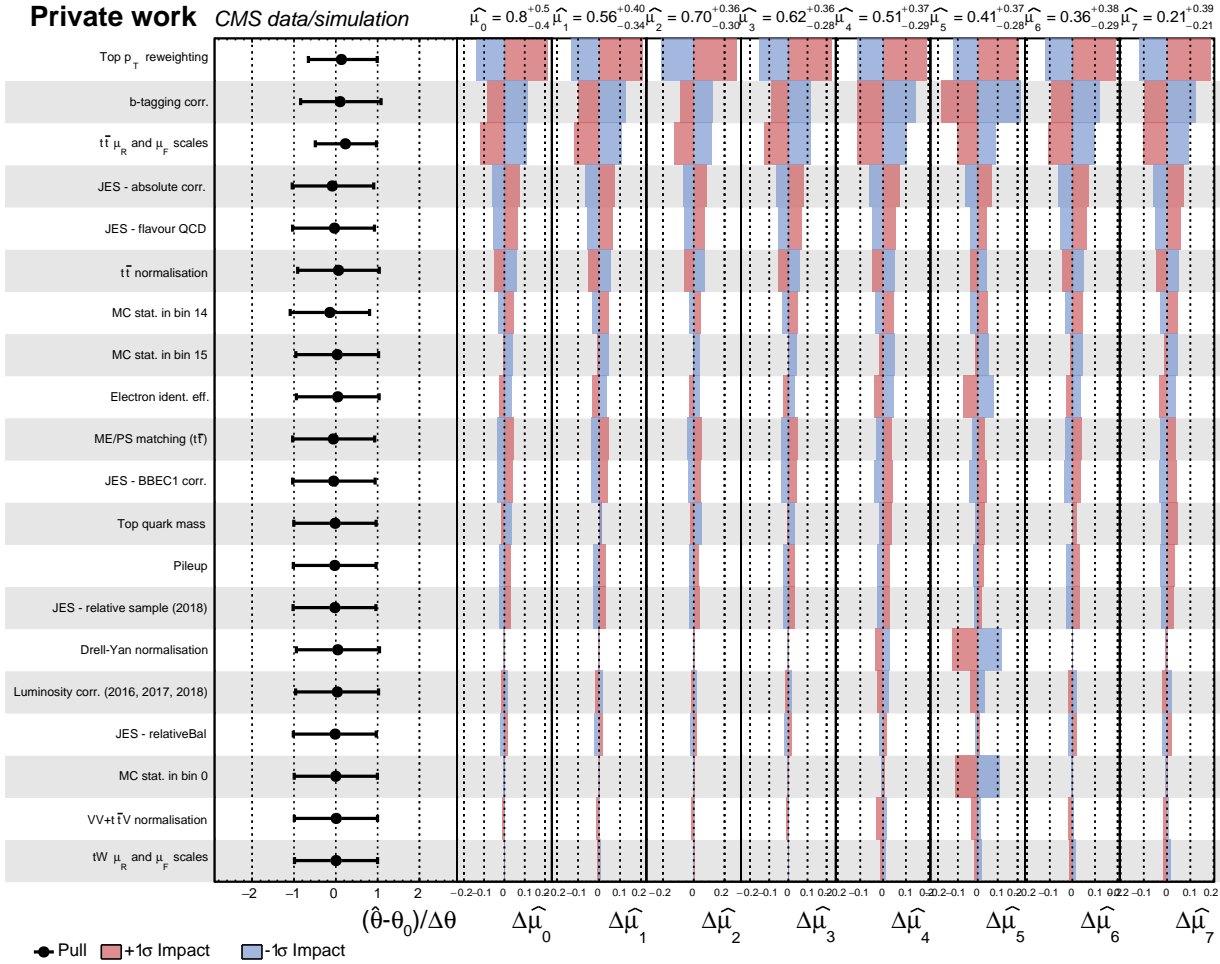
Figure 7.7: The 20 largest impacts (right columns) and pulls (second column) of the nuisance parameters listed in the first column from the fit used to determine the differential tW cross section depending on the leading lepton p_{T}. The horizontal bars on the pulls show the ratio of the uncertainties of the fit result to the previous ones, effectively giving the constraint on the nuisance parameter. The label "corr." refers to the correlated component of the uncertainty over the three years and "uncorr." the uncorrelated component for each year. The entries are defined and detailed in Chap. 5.
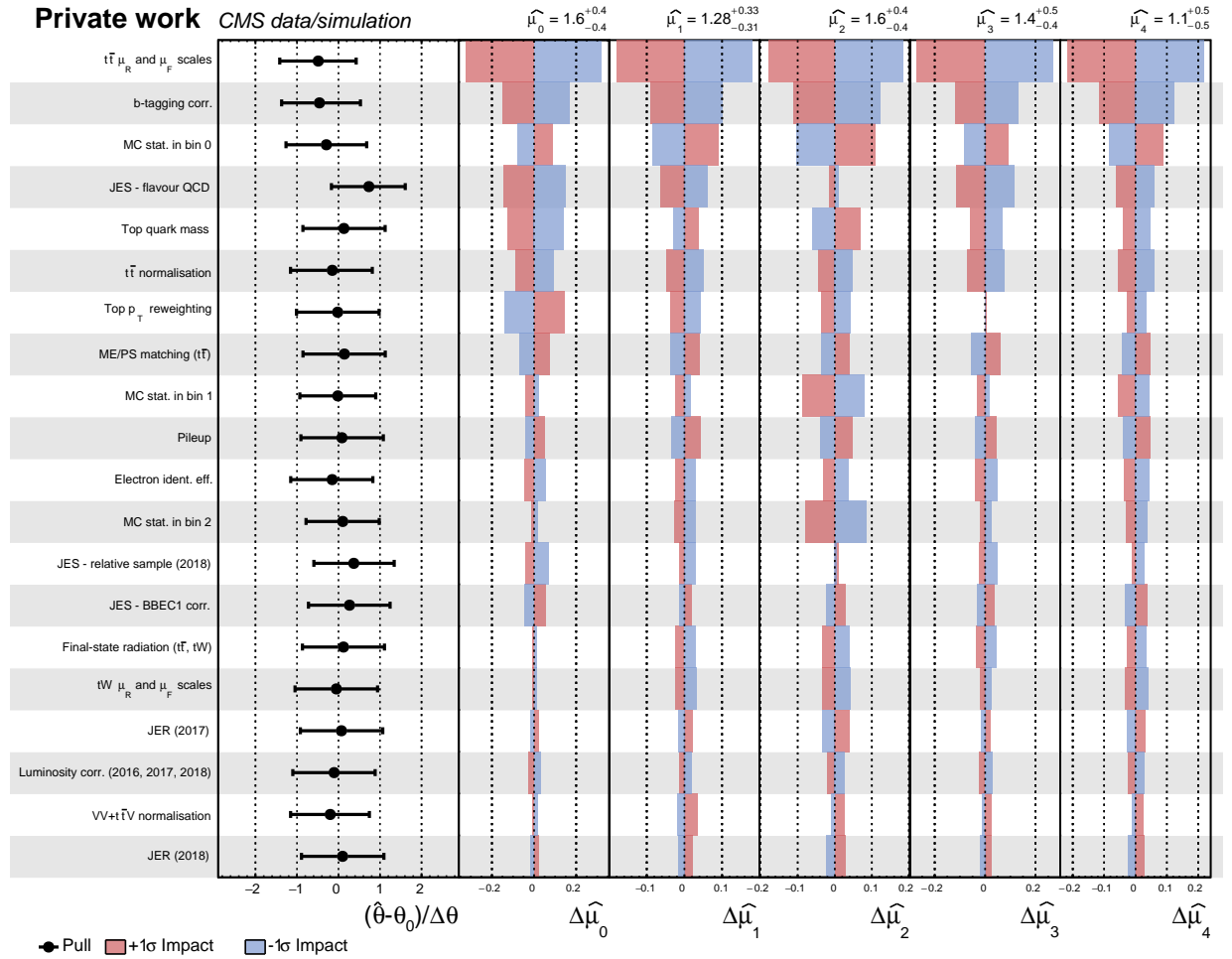
Figure 7.8: The 20 largest impacts (right columns) and pulls (second column) of the nuisance parameters listed in the first column from the fit used to determine the differential tW cross section depending on the jet p_{T}. The horizontal bars on the pulls show the ratio of the uncertainties of the fit result to the previous ones, effectively giving the constraint on the nuisance parameter. The label "corr." refers to the correlated component of the uncertainty over the three years and "uncorr." the uncorrelated component for each year. The entries are defined and detailed in Chap. 5.
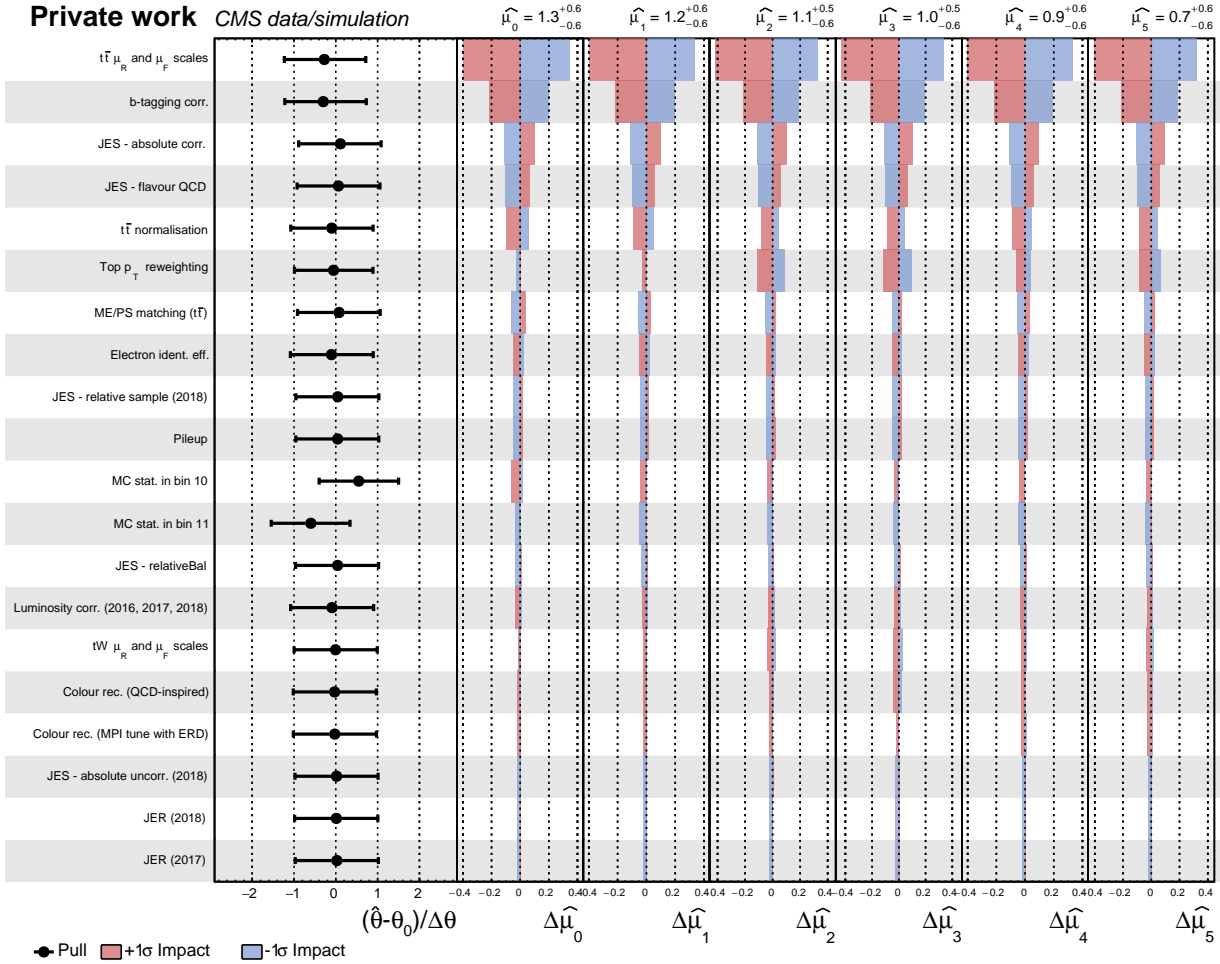
Figure 7.9: The 20 largest impacts (right columns) and pulls (second column) of the nuisance parameters listed in the first column from the fit used to determine the differential tW cross section depending on the $\Delta\varphi(e^{\pm}, \mu^{\mp})$. The horizontal bars on the pulls show the ratio of the uncertainties of the fit result to the previous ones, effectively giving the constraint on the nuisance parameter. The label "corr." refers to the correlated component of the uncertainty over the three years and "uncorr." the uncorrelated component for each year. The entries are defined and detailed in Chap. 5.
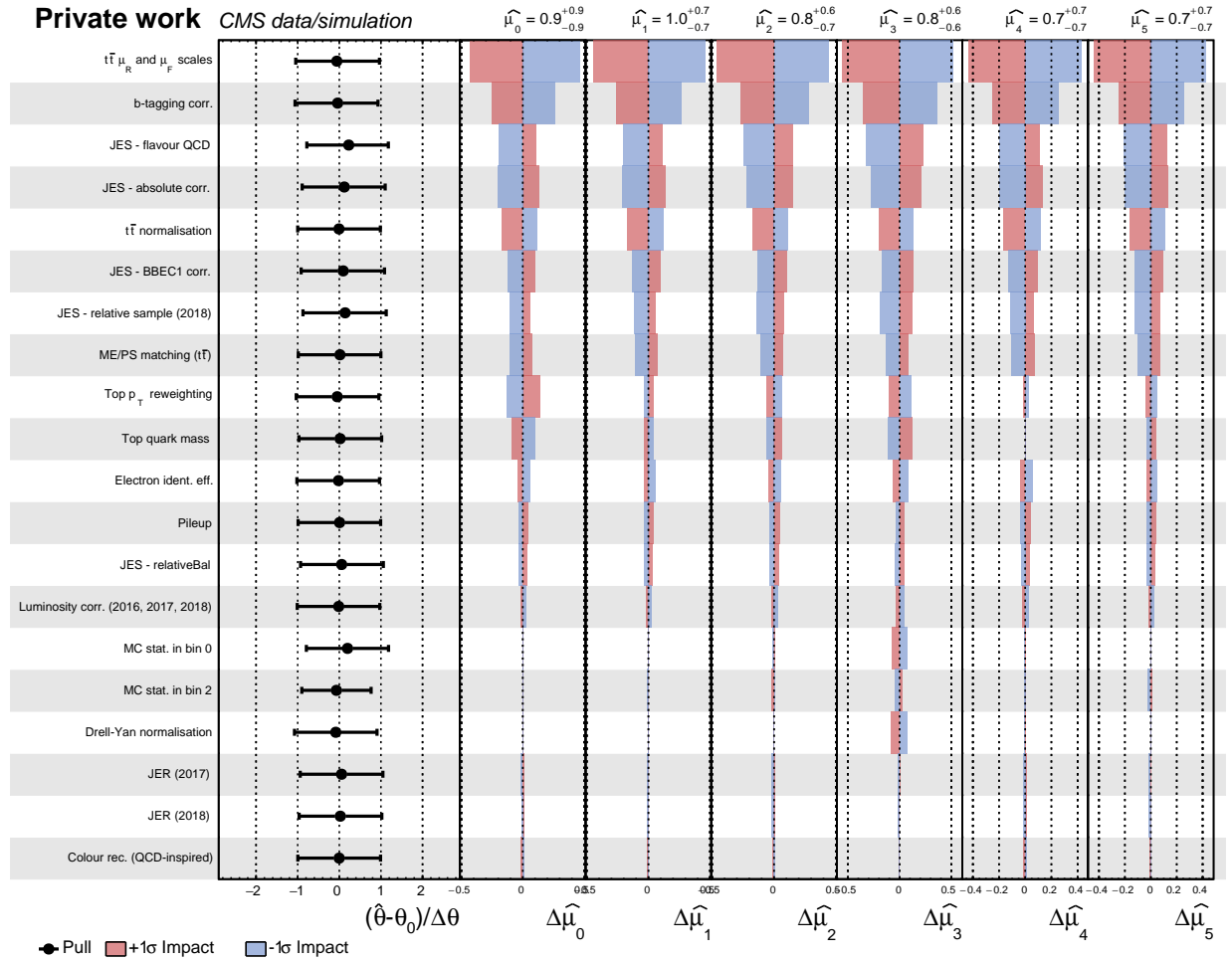
Figure 7.10: The 20 largest impacts (right columns) and pulls (second column) of the nuisance parameters listed in the first column from the fit used to determine the differential tW cross section depending on the $m(e^{\pm}, \mu^{\mp}, \mathrm{j})$. The horizontal bars on the pulls show the ratio of the uncertainties of the fit result to the previous ones, effectively giving the constraint on the nuisance parameter. The label "corr." refers to the correlated component of the uncertainty over the three years and "uncorr." the uncorrelated component for each year. The entries are defined and detailed in Chap. 5.
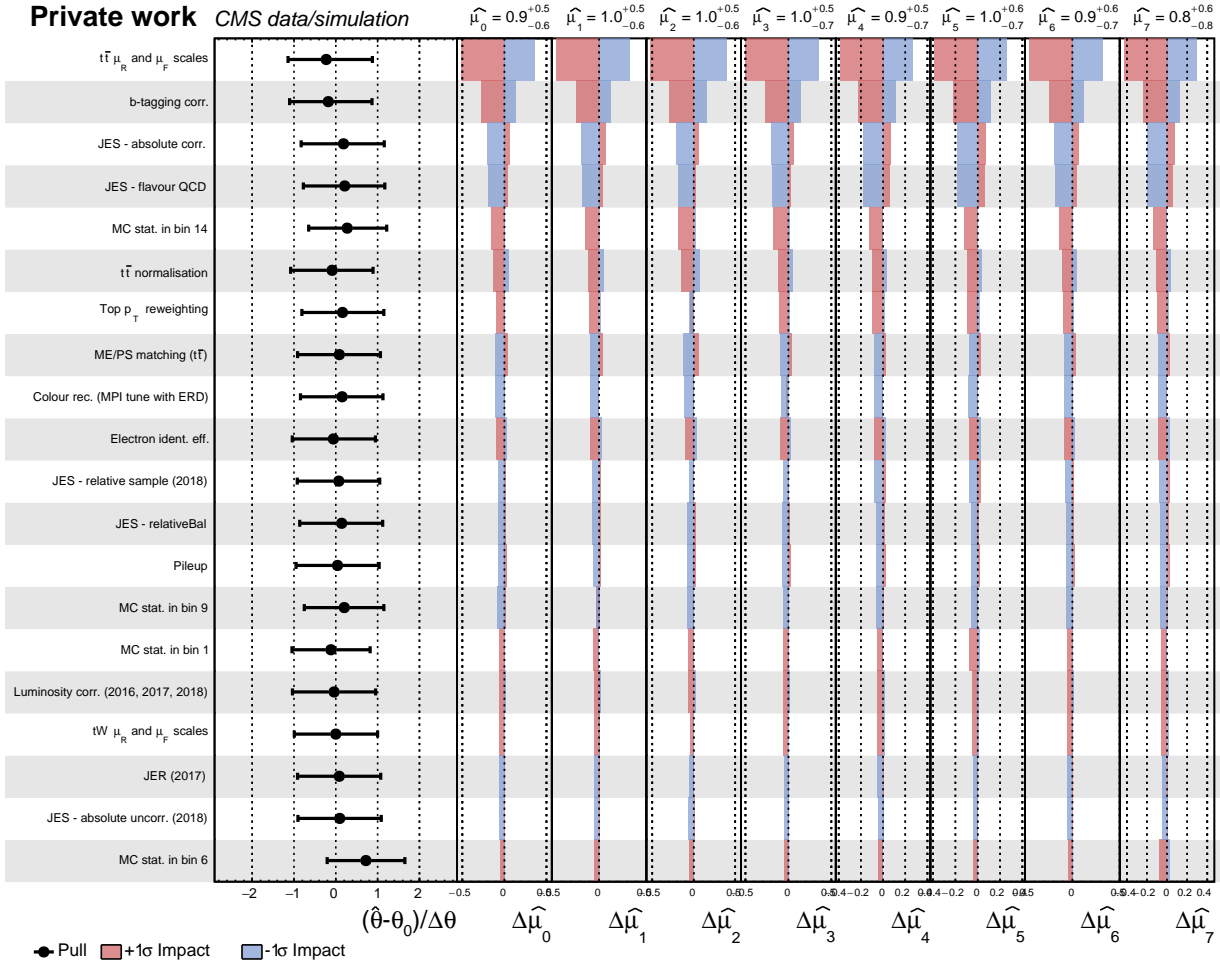
Figure 7.11: The 20 largest impacts (right columns) and pulls (second column) of the nuisance parameters listed in the first column from the fit used to determine the differential tW cross section depending on the $p_z(e^{\pm}, \mu^{\mp}, j)$. The horizontal bars on the pulls show the ratio of the uncertainties of the fit result to the previous ones, effectively giving the constraint on the nuisance parameter. The label "corr." refers to the correlated component of the uncertainty over the three years and "uncorr." the uncorrelated component for each year. The entries are defined and detailed in Chap. 5.
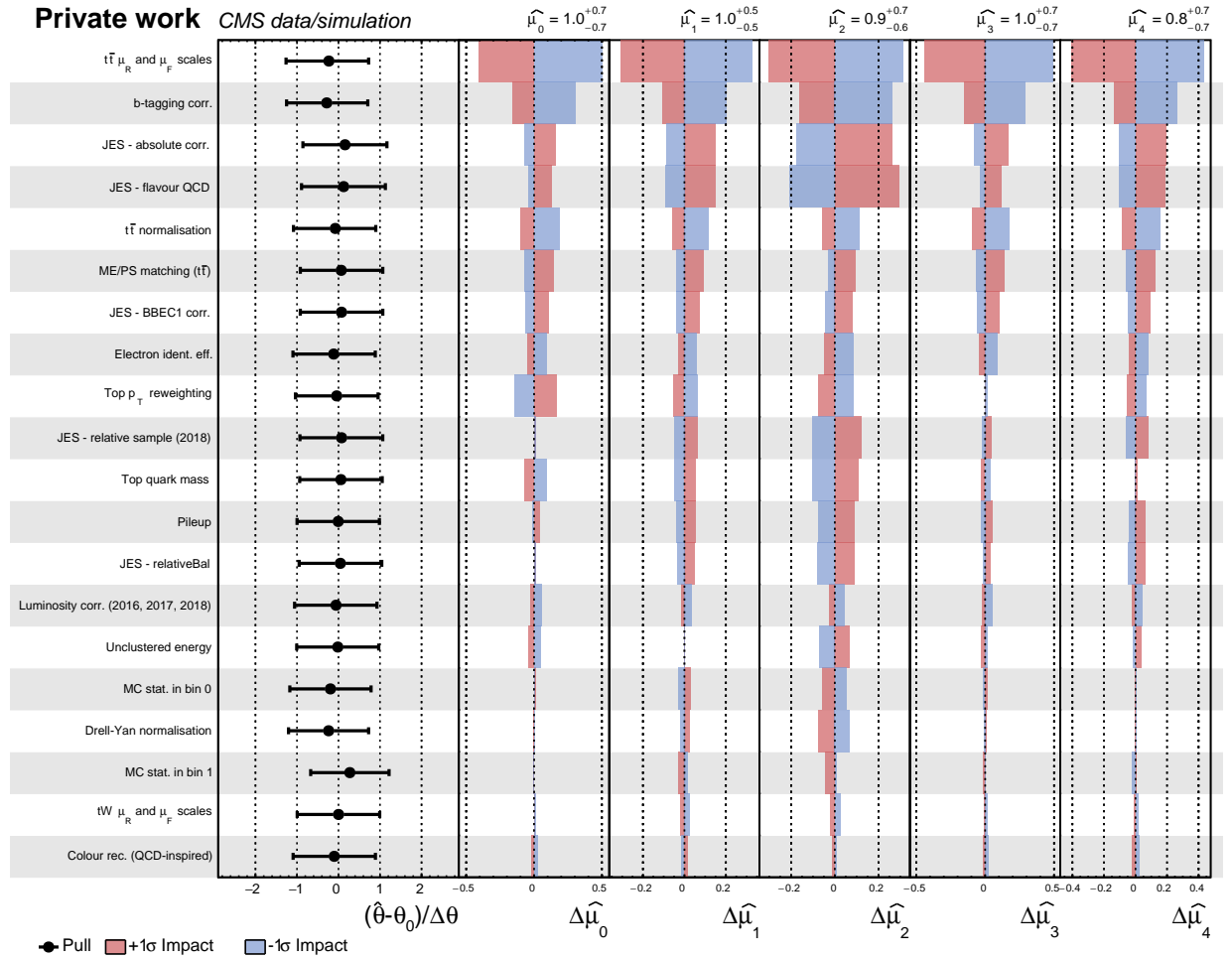
Figure 7.12: The 20 largest impacts (right columns) and pulls (second column) of the nuisance parameters listed in the first column from the fit used to determine the differential tW cross section depending on the $m_{\mathrm{T}}(e^{\pm}, \mu^{\mp}, \mathrm{j}, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$. The horizontal bars on the pulls show the ratio of the uncertainties of the fit result to the previous ones, effectively giving the constraint on the nuisance parameter. The label "corr." refers to the correlated component of the uncertainty over the three years and "uncorr." the uncorrelated component for each year. The entries are defined and detailed in Chap. 5.

8 Summary and conclusions

THE work presented in this thesis uses data collected during the Run 2 at the CMS experiment with proton-proton collisions provided by the LHC at $\sqrt{s} = 13$ TeV of energy in the centre-of-mass frame. A total of 138 fb^{-1} integrated luminosity was exploited to measure the top quark production process in association with a W boson (tW). In addition, research work for the CMS Muon System for Phase-2 of LHC was performed.

The inclusive cross-section of the tW process was measured using dileptonic events in the e$^{\pm}\mu^{\mp}$ channel: they have been categorised depending on the number of jets and the number of jets originating from the fragmentation of bottom quarks. The signal is measured using a maximum likelihood fit, obtaining:

$$\sigma_{\text{tW}}^{\text{obs.}} = 79.2 \pm 0.9(\text{stat.}) \,^{+7.7}_{-8.0}(\text{syst.}) \pm 1.2(\text{lumi.}) \text{ pb.}$$

This result has a total relative uncertainty of 10%. Systematic sources of uncertainty dominate the measurement, being those with largest impact the jet energy scale corrections, the normalisation of the non-W/Z background, the matrix element scales of the tW process, and the modelling of the final-state radiation of the $t\bar{t}$ and tW processes.

The observed value agrees within uncertainties with recent measurements of this process at the same energy in the same channel from ATLAS, as well as from CMS in the semileptonic decay channel. The result is an evolution and optimisation of the observation with only 2016 data from CMS, profiting from the updates and the improved understanding of the simulation modelling, calibrations, corrections, methodology, etc. gained since then. Our new value agrees within two sigma with the previous value and is the most precise measurement so far of the tW inclusive cross-section. Furthermore, it is found to be consistent with the SM predictions at aNNLO in QCD, $\sigma_{\text{tW-ch.}} = 71.7 \pm 1.8\,(\text{scale}) \pm 3.4\,(\text{PDF},\alpha_S) \text{ pb} = 71.7 \pm 3.8$ pb, and with the latest at aN^{3}LO, also in QCD, $\sigma_{\text{tW-ch.}} = 79.3^{+1.9}_{-1.8}\,(\text{scale}) \pm 2.2\,(\text{PDF},\alpha_S) \text{ pb} = 79.3^{+2.9}_{-2.8}$ pb. A summary of all tW measurements at a centre-of-mass energy of $\sqrt{s} = 13$ TeV with proton-proton collisions can be seen in Fig. 8.1, and a larger comparison with the other single top process in Fig. 8.2.

Normalised differential cross-section measurements of this production mode have been done depending on six kinematic observables of the final-state objects. They are unfolded to particle-level, using a fiducial region similar to the signal region. The results are confronted with different simulation generators at NLO in QCD interfaced with two different parton showering algorithms. These results have relative uncertainties in the range of 10–50%, depending on the measured
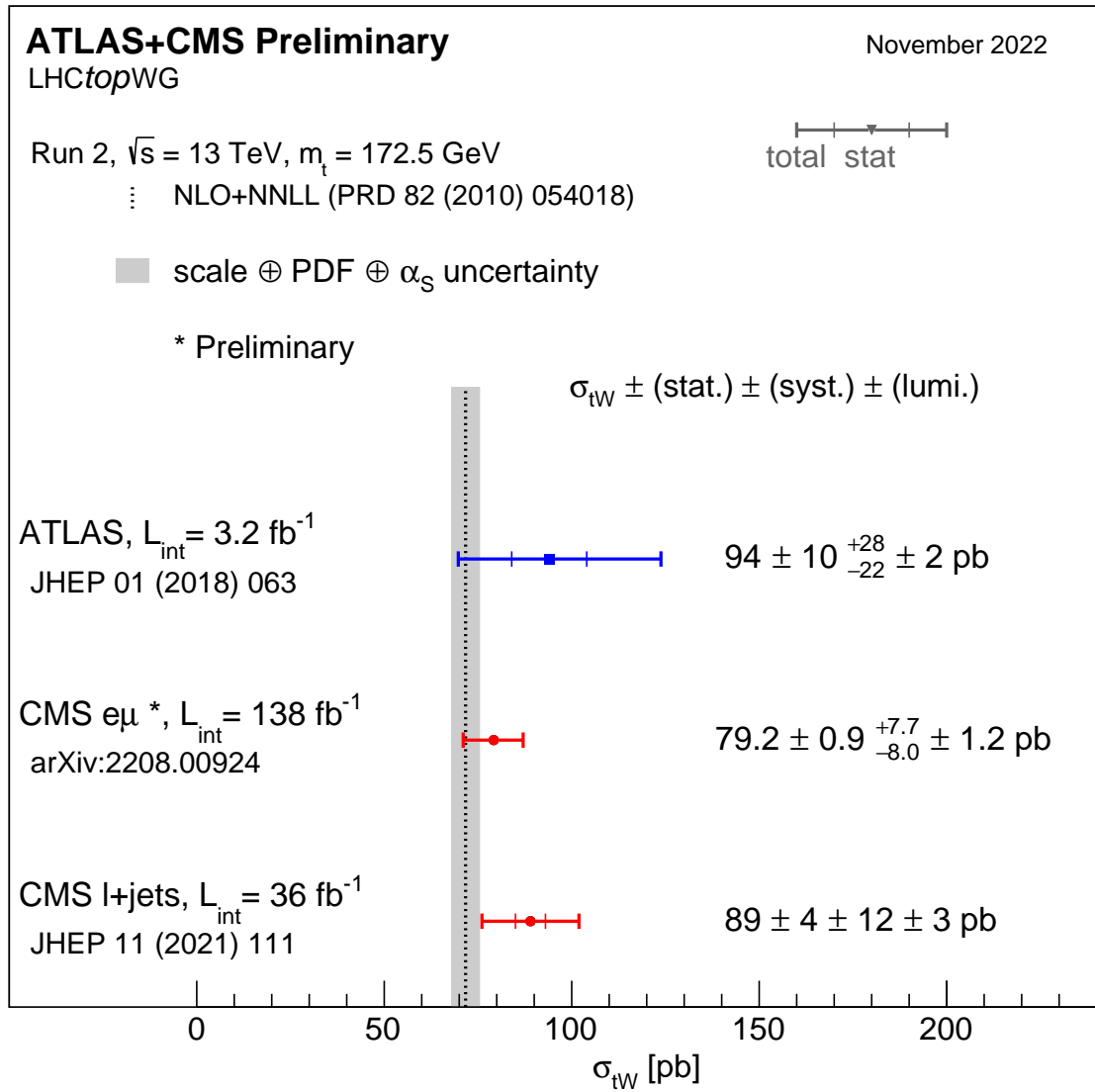
Figure 8.1: Inclusive cross section measurements of the tW process at a centre-of-mass energy of $\sqrt{s} = 13$ TeV with proton-proton collisions in both dileptonic as well as semileptonic channels.

ATLAS+CMS Preliminary
LHC*top*WG

Single top-quark production
November 2022

Inclusive cross-section [pb]

t-channel
tW
s-channel

t-channel
ATLAS PRD 90 (2014) 112006, EPJC 77 (2017) 531, JHEP 04 (2017) 086
CMS JHEP 12 (2012) 035, JHEP 06 (2014) 090, PLB 800 (2019) 135042
LHC comb. JHEP 05 (2019) 088

tW
ATLAS PLB 716 (2012) 142, JHEP 01 (2016) 064, JHEP 01 (2018) 063
CMS PRL 110 (2013) 022003, PRL 112 (2014) 231802, arXiv:2208.00924
LHC comb. JHEP 05 (2019) 088

s-channel
ATLAS PLB 756 (2016) 228, arXiv:2209.08990
CMS JHEP 09 (2016) 027
LHC comb. JHEP 05 (2019) 088

NNLO PLB 736 (2014) 58
scale uncertainty

NLO + NNLL PRD 83 (2011) 091503,
PRD 82 (2010) 054018, PRD 81 (2010) 054028
tW: t$\bar{t}$ contribution removed
scale ⊕ PDF ⊕ α_s uncertainty

NLO NPPS 205 (2010) 10, CPC 191 (2015) 74
$\mu_R = \mu_F = m_{top}$,
CT10nlo, MSTW2008nlo, NNPDF2.3nlo
tW: p_T^b veto for t$\bar{t}$ removal = 60 GeV and μ_F = 65 GeV
scale uncertainty
scale ⊕ PDF ⊕ α_s uncertainty

stat. total

$\sqrt{s}$ [TeV]

Figure 8.2: Summary of the single top production processes cross sections measured at a centre-of-mass energy of $\sqrt{s} = 13$ TeV with proton-proton collisions.

observable, with larger values in the tails of the distributions. Monte Carlo modelling is in general the main source of uncertainty, closely followed by the experimental ones.

There is an overall agreement with the SM expectations. There are, however, some discrepancies that have been also observed in other top quark production modes. The different approaches used in the differential measurement to simulate the tW events give similar values in all the distributions, which points to small effects due to the interference of the tW and t$\bar{t}$ processes on these distributions in the defined fiducial region. Comparison with the differential results with the ATLAS analysis at the same energy and channel is hard because of differences in the binning and the uncertainties considered. However, both measurements are dominated by systematic experimental and modelling uncertainty sources.

Future prospects in measuring this process should push the frontier of precision beyond what this thesis does. Either at $\sqrt{s} = 13$ TeV or at $\sqrt{s} = 13.6$ TeV in the Run 3 of the LHC, there are various elements that could help in improving the measurement, derivable from the discussions of Chapters 6 and 7. The most important sources of uncertainty in both measurements are either experimental or related to our understanding of the MC modelling. Therefore, improvements in the estimation of both would directly benefit the results. A better result may come from adding sources that estimate the uncertainty in misidentified leptons from jets as final-state leptons (such as the modelling of b quark fragmentation or the semileptonic B meson decay branching ratios) that would help reducing the normalisation uncertainty of the Non-W/Z background, the second

source with largest impact in the signal extraction of the inclusive measurement. Given the close relationship between the tW and t$\bar{\text{t}}$ processes (that interfere at NLO in QCD), and the impact the latter has as background on any measurement done of the former, a combined measurement of both processes could improve the results. Another alternative would be to directly measure the WWbb process, that includes both tW and t$\bar{\text{t}}$ diagrams, and would allow to remove the uncertainty due to the interference removal methods (DR/DS).

New differential studies might attempt to increase the resolution of the observables (by adding more bins), or the range considered. Such endeavours may find very difficult to stabilise the maximum likelihood fit, as its complexity escalates largely with the number of bins considered. To overcome this, an alternative approach that separates the signal extraction step and the unfolding (such as the used in [2]) could be helpful, probably providing a more robust and easier to understand (although not as powerful, a priori) solution.

The research work developed for the Phase-2 of LHC upgrade of the CMS detector is also presented in this thesis. This will prepare the experiment for the HL-LHC, that will provide the detectors with a boost in the luminosity achieved, and thus in the collected data. This also poses dangers to the detectors due to the increase in radiation. Longevity studies of the DT subdetectors of CMS have been done by radiating a spare muon chamber in the GIF++ facility at CERN. This experiment allowed us to estimate the effect of the high radiation doses expected during HL-LHC on the hit efficiency of the chambers. When this effect was simulated in the CMS reconstruction software, the muon reconstruction efficiency showed an overall small variation. No dependency on the muon p_T or φ was observed. However, a slight dependency on η is shown, with the largest reduction in efficiency, of the order of 1%, happening only in small ranges of η, related to the barrel-endcap transition region. This values were obtained for an scenario where the chambers accumulated a radiation dose equivalent to two times the expected integrated luminosity at the end of the HL-LHC.

Additional studies cover the effect of ageing in the DT L1 trigger and its evolution thanks to the upgrade of the electronic hardware that will be used. Two algorithms, the so-called analytical method (AM) and the Hough transform-based algorithm (HB) were put to the test using MC samples simulated with the same average simultaneous collisions expected in the HL-LHC ($\mu =$ 200). A reduction in trigger efficiency is expected especially in the most inner (MB1s) and external (MB4s) muon chambers. Those chambers with largest η with respect to the interaction point show a higher reduction on trigger efficiency. When considering the same ageing scenario mentioned above, a drop in the muon trigger efficiency was observed of the order of 30% for the AM and 50% for the HB algorithms in the worst cases of the most inner chambers. In the intermediate chambers the decrease in efficiency was well below 5%, while in the external DTs is expected to be smaller than 10%. Globally the AM algorithm performs better than the HB. In addition, the former can recover up to a 20% of the lost efficiency by using RPC detector information. This can be extended to the spatial and temporal resolutions, that were also studied.

9 Resumen y conclusiones

E L trabajo presentado en esta tesis utilizada datos provenientes del experimento CMS recogidos durante el *Run* 2del LHC en colisiones protón-protón a una energía en el centro de masas de $\sqrt{s} = 13$ TeV. Se usó una luminosidad integrada total de 138 fb^{-1} fue usada para medir el proceso de producción del *quark top* en asociación con un bosón W (tW). Además, se realizó trabajo de investigación para el Sistema de Muones de CMS de cara a la Fase-2 del LHC.

La sección eficaz inclusiva del proceso tW fue medida usando sucesos dileptónicos en el canal e$^{\pm}\mu^{\mp}$: estos han sido categorizados dependiendo del número de *jets* observado, así como la cantidad de aquellos que provienen de la fragmentación de *quarks bottom*. La señal se ha medido usando un ajuste de máxima verosimilitud, obteniendo:

$$\sigma_{\text{tW}}^{\text{obs.}} = 79.2 \pm 0.9(\text{estad.}) \, {}^{+7.7}_{-8.0}(\text{sist.}) \pm 1.2(\text{lumi.}) \ \text{pb}.$$

Este resultado tiene una incertidumbre total relativa del 10%. Las fuentes sistemáticas de incertidumbre dominan la medida, siendo aquellas con mayor impacto las correcciones de la energía de los *jets*, la normalización en el fondo no-W/Z, las escalas de energía del elemento de matriz del proceso tW, y la modelización de la radiación del estado final de los procesos t$\bar{\text{t}}$ y tW.

El valor observado está de acuerdo, dentro de las incertidumbres, con las medidas recientes de este proceso a la misma energía y en el mismo canal hechas por ATLAS, así como con la de CMS en el canal de desintegración semileptónico. El resultado es una evolución y optimización de la observación realizada en 2016 por CMS, beneficiándose de las actualizaciones y el incremento en la comprensión de la modelización de las simulaciones, calibraciones, correcciones, metodología, etcétera, que se ha conseguido desde entonces. Además, la medida es consistente con las predicciones del SM a aNNLO en QCD, $\sigma_{\text{canal tW}} = 71.7 \pm 1.8 \, (\text{escalas}) \pm 3.4 \, (\text{PDF}, \alpha_S) \ \text{pb} = 71.7 \pm 3.8 \ \text{pb}$, y con las más recientes a aN^{3}LO, también en QCD, $\sigma_{\text{canal tW}} = 79.3^{+1.9}_{-1.8} \, (\text{escalas}) \pm 2.2 \, (\text{PDF}, \alpha_S) \ \text{pb} = 79.3^{+2.9}_{-2.8} \ \text{pb}$. Se puede observar en la figura 9.1 un resumen de todas las medidas del proceso tW a una energía en centro de masas de $\sqrt{s} = 13$ TeV con colisiones protón-protón, y de forma más general, con otros procesos de producción individual de *quarks top*, en la figura 9.2.

Las medidas de la sección eficaz diferencial normalizada de este modo de producción se han realizado dependiendo de seis observables cinemáticos de los objetos del estado final. Ellos son deconvolucionados a nivel de partícula, usando una región fiducial similar a la región de señal. Los resultados son contrastados con varios generadores de simulaciones a NLO en QCD, articuladas
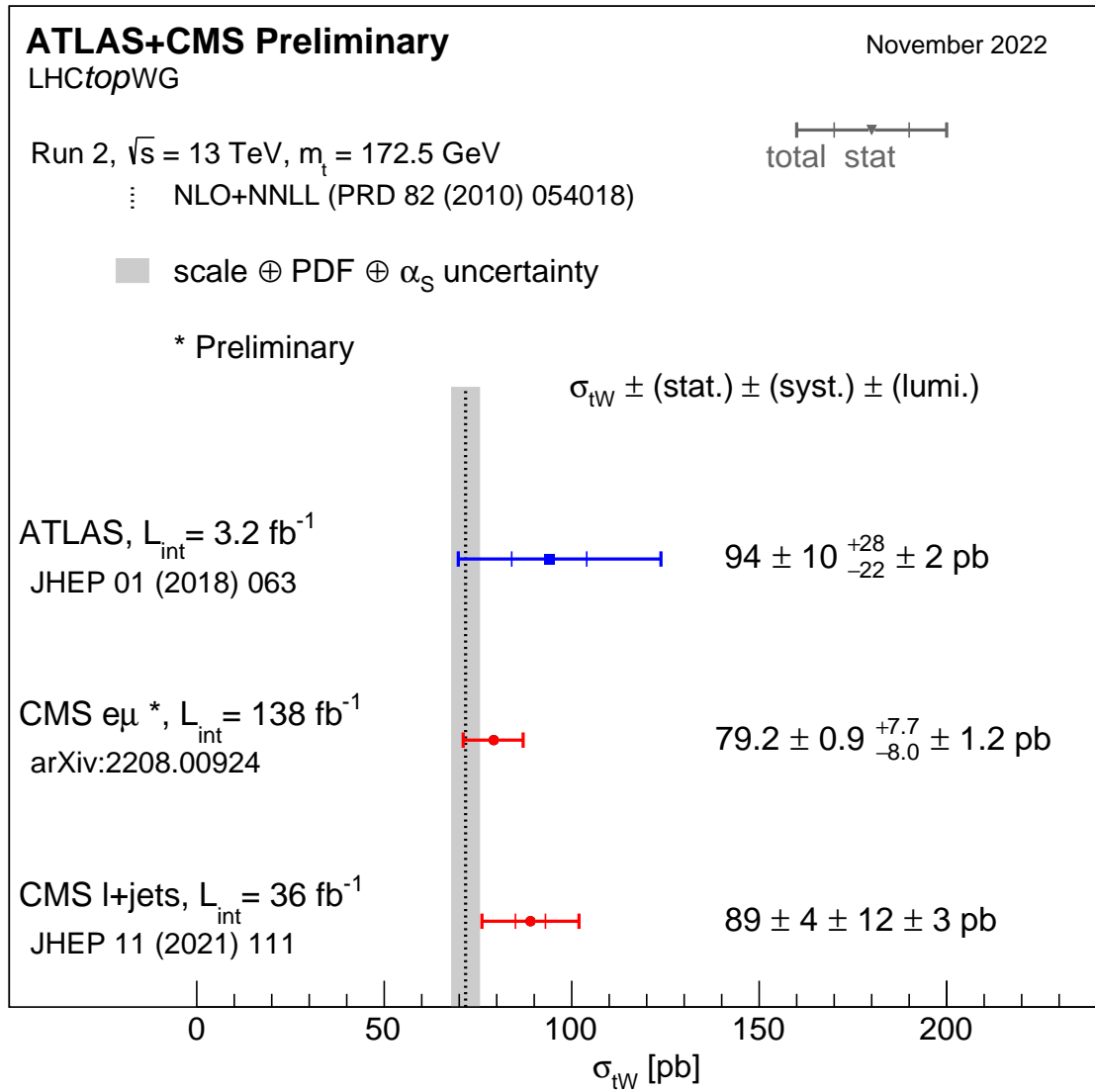
Figure 9.1: Medidas de la sección eficaz inclusiva del proceso tW a una energía en centro de masas de $\sqrt{s} = 13$ TeV, en colisiones protón-protón en los canales dileptónico y semileptónico.
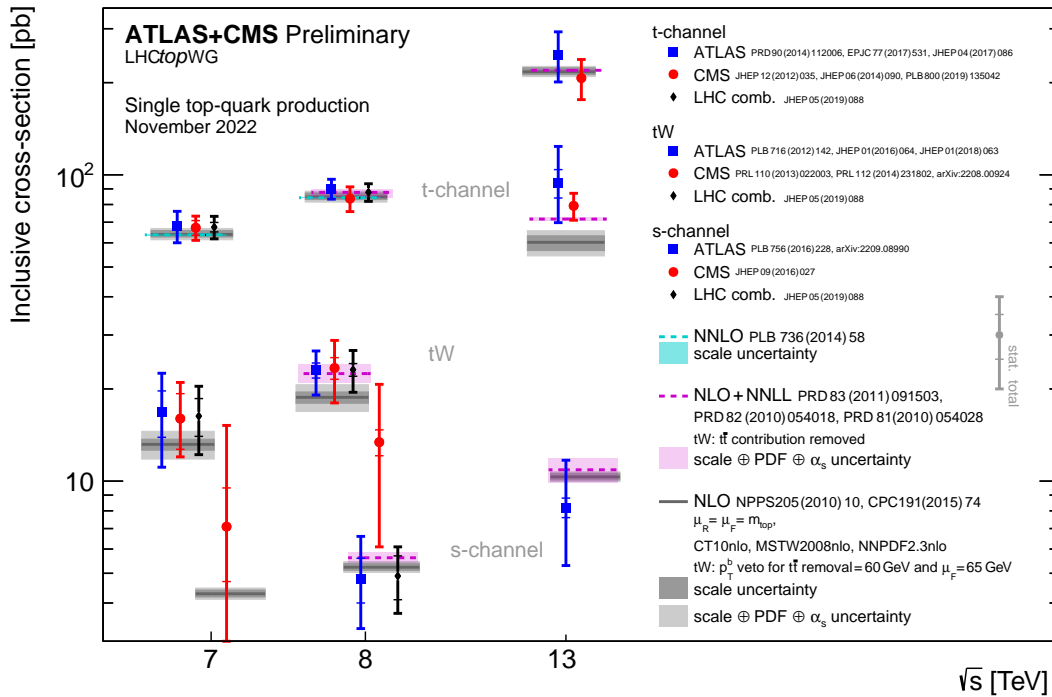
Figure 9.2: Resumen de las secciones eficaces de producción individual de *quarks top* medidas a una energía en centro de masas de $\sqrt{s} = 13$ TeV en colisiones protón-protón.

con dos algoritmos distintos de simulación de cascadas de partones. Estos resultados tienen incertidumbres relativas en el intervalo de 10–50%, dependiendo del observable medido, con valores mayores en las colas de las distribuciones. La modelización del MC es globalmente la principal fuente de incertidumbre, seguida de las fuentes experimentales.

Los resultados son globalmente compatibles con las expectativas del SM. Hay, sin embargo, algunas discrepancias que también han sido observadas en otros modos de producción de *quarks top*. Los distintos generadores para simular sucesos de tW mostrados en los resultados diferenciales ofrecen valores similares en todas las distribuciones, lo que apunta a pequeños efectos debido a la interferencia de los procesos tW y $t\bar{t}$ en estas variables y en la región fiducial escogida. La comparación de los resultados diferenciales con el análisis de ATLAS a la misma energía y en el mismo canal de desintegración es difícil, debido a las diferencias en el ancho y número de los *bins* y a las incertidumbres consideradas. Sin embargo, ambos resultado están dominados por las fuentes de incertidumbre experimentales y de modelización de MC.

Las perspectivas futuras en la medida de este proceso deberían mejorar la precisión más allá de lo que lo hace esta tesis. Sea a $\sqrt{s} = 13$ TeV, o a $\sqrt{s} = 13.6$ TeV en el *Run* 3, hay varios elementos que podrían ayudar en la mejora de la medida, que se pueden derivar de las discusiones de los capítulos 6 y 7. Las fuentes de incertidumbre más importantes en ambas medidas son bien experimentales o relacionadas con el entendimiento de la modelización del MC. Las mejoras en la estimación de ambas afectarían directamente a los resultados. Una mejora en la precisión podría

conseguirse también incluyendo fuentes de incertidumbre dedicadas a la correcta modelización de aquellos procesos en los que se identifican erróneamente leptones procedentes de *jets* como si proviniesen del estado final (como, por ejemplo, la modelización de la fragmentación del *quark* b, o las fracciones de desintegración semileptónica de los mesones B). Esto ayudaría indirectamente a reducir la incertidumbre de normalización del grupo de fondos no-W/Z (que constituye el segundo mayor impacto en la extracción de señal de la medida inclusiva). Dada la relación cercana entre los procesos tW y t$\bar{\text{t}}$ (que interfieren a NLO en QCD), y el impacto que el segundo tiene como contribución de fondo del primero, una medida combinada de ambos modos de producción podría mejorar los resultados. Otra alternativa podría ser hacer directamente una medida del proceso WWbb, que incluye los diagramas de tW y t$\bar{\text{t}}$, y que permitiría eliminar de raíz la incertidumbre en el análisis debido a los métodos de eliminación de la interferencia en sucesos de tW (DR/DS).

Nuevos estudios diferenciales podrían intentar aumentar la resolución de los observables (agregando un mayor número de *bins*), o el rango considerado en los mismos. Tales esfuerzos podrían encontrar muy difícil la estabilización del ajuste de máxima verosimilitud, puesto que su complejidad escala significativamente con el número de *bins* considerado. Para evitar esta problemática, una solución podría ser separar la extracción de señal y la deconvolución (tal y como se hace en [2]), ofreciendo probablemente una metodología más robusta y fácilmente entendible (si bien, *a priori*, menos potente).

El trabajo realizado para la actualización de la Fase-2 del LHC del detector CMS también se muestra en esta tesis. Estos esfuerzos prepararán el experimento para el HL-LHC, que ofrecerá a los detectores un incremento sustancial en la luminosidad observada, y así indirectamente, en la cantidad de datos medidos. La actualización también conlleva peligros para los detectores debidos al incremento en la radiación a la que serán sometidos. Se han realizado estudios de longevidad de los subdetectores DT de muones de CMS, radiando para ello una cámara de muones de repuesto en la instalación GIF++ del CERN. Este experimento nos permite estimar el efecto de las altas dosis de radiación esperadas con el HL-LHC en la eficiencia de señal (*hit*) de las cámaras. Al simular este efecto en los programas de reconstrucción de CMS, la eficiencia de reconstrucción de muones mostró en general una variación pequeña. No se encontró dependencia en función del p_T o φ del muon. Sin embargo, sí se observó una leve dependencia en η, donde la mayor reducción en eficiencia (del orden del 1%) aparece solo en pequeñas regiones de η, relacionadas con la zona de transición entre el barril y las tapas del detector. Estos valores se obtuvieron para un escenario de envejecimiento en el que las cámaras acumulan una dosis de radiación equivalente a dos veces la que se espera al final del HL-LHC.

Se hicieron estudios adicionales para abordar el efecto del envejecimiento de los detectores DT sobre el sistema L1 de *trigger* y su posible evolución aprovechando la actualización de la electrónica que se usará. Se pusieron a prueba dos algoritmos, el llamado método analítico (AM) y el basado en la transformada de Hough (HB), empleando muestras de MC simuladas con la misma cantidad de simulaciones simultáneas esperadas en el HL-LHC ($\mu = 200$). Se espera una reducción en la

eficiencia de *trigger* en las cámaras de muones más internas (MB1s) y externas (MB4s). Aquellas cámaras de muones con una mayor inclinación en η con respecto al punto de interacción de CMS mostraron una reducción mayor en la eficiencia de *trigger*. Al considerar el mismo escenario de envejecimiento descrito en el párrafo anterior se observa un descenso en la eficiencia de *trigger* de muones del orden del 30% en el caso del algoritmo AM y del 50% para HB en las simulaciones en los peores casos de las cámaras más internas. En las cámaras intermedias, el descenso en esta eficiencia era menor del 5%, mientras que en los subdetectores DT externos se espera que sea de menos del 10%. Globalmente, el algoritmo AM ofrece mejor rendimiento que el HB. Además, el primero puede recuperar hasta un 20% de la eficiencia perdida al incorporar información de los subdetectores RPC. Estas conclusiones globales se pueden extender a las resoluciones espacial y temporal, que también fueron estudiadas.

Bibliography

[1] "Measurement of inclusive and differential cross sections for single top quark production in association with a W boson in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: (Aug. 2022). arXiv: 2208.00924 [hep-ex].

[2] *Measurement of differential cross sections for single top quark production in association with a W boson at $\sqrt{s} = 13$ TeV*. Tech. rep. Geneva: CERN, 2020. URL: https://cds.cern.ch/record/2712818.

[3] Albert M Sirunyan et al. "Measurement of the production cross section for single top quarks in association with W bosons in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *JHEP* 10 (2018), p. 117. DOI: 10.1007/JHEP10(2018)117. arXiv: 1805.07399 [hep-ex].

[4] Wikimedia Commons. *Standard Model of Elementary Particles*. File: Standard_Model_of_Elementary_Particles.svg. 2019. URL: https://commons.wikimedia.org/wiki/File:Standard%5C_Model%5C_of%5C_Elementary%5C_Particles.svg.

[5] S. L. Glashow. "Partial Symmetries of Weak Interactions". In: *Nucl. Phys.* 22 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2.

[6] Steven Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264. URL: https://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[7] A. Salam and J.C. Ward. "Electromagnetic and weak interactions". In: *Physics Letters* 13.2 (1964), pp. 168–171. ISSN: 0031-9163. DOI: https://doi.org/10.1016/0031-9163(64)90711-5. URL: https://www.sciencedirect.com/science/article/pii/0031916364907115.

[8] C. S. Wu et al. "Experimental Test of Parity Conservation in Beta Decay". In: *Phys. Rev.* 105 (4 Feb. 1957), pp. 1413–1415. DOI: 10.1103/PhysRev.105.1413. URL: https://link.aps.org/doi/10.1103/PhysRev.105.1413.

[9] M. Gell-Mann. "The interpretation of the new particles as displaced charge multiplets". In: *Il Nuovo Cimento (1955-1965)* 4.2 (Apr. 1956), pp. 848–866. ISSN: 1827-6121. DOI: 10.1007/BF02748000. URL: https://doi.org/10.1007/BF02748000.

[10] Tadao Nakano and Kazuhiko Nishijima. "Charge Independence for V-particles*". In: *Progress of Theoretical Physics* 10.5 (Nov. 1953), pp. 581–582. ISSN: 0033-068X. DOI: 10.1143/PTP.10.581. eprint: https://academic.oup.com/ptp/article-pdf/10/5/581/5364926/10-5-581.pdf. URL: https://doi.org/10.1143/PTP.10.581.

[11] F. Close. *The Infinity Puzzle: The Personalities, Politics, and Extraordinary Science Behind the Higgs Boson*. OUP Oxford, 2013. ISBN: 9780199673308. URL: https://books.google.bi/books?id=Po1ijAJEvqOC.

[12] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.321.

[13] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.508.

[14] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. "Global Conservation Laws and Massless Particles". In: *Phys. Rev. Lett.* 13 (20 Nov. 1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.585.

[15] H. Fritzsch, M. Gell-Mann, and H. Leutwyler. "Advantages of the color octet gluon picture". In: *Physics Letters B* 47.4 (1973), pp. 365–368. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(73)90625-4. URL: https://www.sciencedirect.com/science/article/pii/0370269373906254.

[16] C. N. Yang and R. L. Mills. "Conservation of Isotopic Spin and Isotopic Gauge Invariance". In: *Phys. Rev.* 96 (1 Oct. 1954), pp. 191–195. DOI: 10.1103/PhysRev.96.191. URL: https://link.aps.org/doi/10.1103/PhysRev.96.191.

[17] Wikimedia Commons. *Brandon Sanderson*. URL: https://en.wikipedia.org/wiki/Brandon_Sanderson (visited on Nov. 29, 2022).

[18] Wikimedia Commons. *File:Running coupling constants.svg — Wikimedia Commons, the free media repository*. [Online; accessed 2023-01-06]. 2020. URL: https://commons.wikimedia.org/w/index.php?title=File:Running_coupling_constants.svg&oldid=507424081.

[19] Raymond Davis, Don S. Harmer, and Kenneth C. Hoffman. "Search for Neutrinos from the Sun". In: *Phys. Rev. Lett.* 20 (21 May 1968), pp. 1205–1209. DOI: 10.1103/PhysRevLett.20.1205. URL: https://link.aps.org/doi/10.1103/PhysRevLett.20.1205.

[20] Q. R. Ahmad et al. "Measurement of the Rate of $\nu_e + d \to p + p + e^-$ Interactions Produced by ^8B Solar Neutrinos at the Sudbury Neutrino Observatory". In: *Phys. Rev. Lett.* 87 (7 July 2001), p. 071301. DOI: 10.1103/PhysRevLett.87.071301. URL: https://link.aps.org/doi/10.1103/PhysRevLett.87.071301.

[21]   William Thomson Baron Kelvin. *Baltimore Lectures on Molecular Dynamics and the Wave Theory of Light.* Cambridge Library Collection - Physical Sciences. Cambridge University Press, 2010. DOI: `10.1017/CBO9780511694523`.

[22]   V. C. Rubin, Jr. Ford W. K., and N. Thonnard. "Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 (R=4kpc) to UGC 2885 (R=122kpc)." In: *ApJ* 238 (June 1980), pp. 471–487. DOI: `10.1086/158003`.

[23]   Vera C. Rubin and Jr. Ford W. Kent. "Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions". In: *ApJ* 159 (Feb. 1970), p. 379. DOI: `10.1086/150317`.

[24]   Adam G. Riess et al. "Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant". In: *The Astronomical Journal* 116.3 (Sept. 1998), pp. 1009–1038. DOI: `10.1086/300499`. URL: `https://doi.org/10.1086/300499`.

[25]   S. Perlmutter et al. "Measurements of Ω and Λ from 42 High-Redshift Supernovae". In: *The Astrophysical Journal* 517.2 (June 1999), pp. 565–586. DOI: `10.1086/307221`. URL: `https://doi.org/10.1086/307221`.

[26]   Tsutomu Yanagida. "Horizontal gauge symmetry and masses of neutrinos". In: *Conf. Proc. C* 7902131 (1979). Ed. by Osamu Sawada and Akio Sugamoto, pp. 95–99.

[27]   Hironari Miyazawa. "Baryon Number Changing Currents*". In: *Progress of Theoretical Physics* 36.6 (Dec. 1966), pp. 1266–1276. ISSN: 0033-068X. DOI: `10.1143/PTP.36.1266`. eprint: `https://academic.oup.com/ptp/article-pdf/36/6/1266/5156437/36-6-1266.pdf`. URL: `https://doi.org/10.1143/PTP.36.1266`.

[28]   Pierre Simon Laplace. *Essai philosophique sur les probabilités.* 5th ed. Cambridge Library Collection - Mathematics. Cambridge University Press, 2009. DOI: `\linebreak10.1017/CBO9780511693182`.

[29]   Christian Bierlich et al. "A comprehensive guide to the physics and usage of PYTHIA 8.3". In: (Mar. 2022). arXiv: `2203.11601 [hep-ph]`.

[30]   Albert M Sirunyan et al. "Measurement of differential cross sections and charge ratios for t-channel single top quark production in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *Eur. Phys. J. C* 80.5 (2020), p. 370. DOI: `10.1140/epjc/s10052-020-7858-1`. arXiv: `1907.08330 [hep-ex]`.

[31]   R. L. Workman et al. "Review of Particle Physics". In: *PTEP* 2022 (2022), p. 083C01. DOI: `10.1093/ptep/ptac097`.

[32]   Georges Aad et al. "Search for a scalar partner of the top quark in the all-hadronic $t\bar{t}$ plus missing transverse momentum final state at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Eur. Phys. J. C* 80.8 (2020), p. 737. DOI: `10.1140/epjc/s10052-020-8102-8`. arXiv: `2004.14060 [hep-ex]`.

[33] Georges Aad et al. "Search for squarks and gluinos in final states with same-sign leptons and jets using 139 fb^{-1} of data collected with the ATLAS detector". In: *JHEP* 06 (2020), p. 046. DOI: 10.1007/JHEP06(2020)046. arXiv: 1909.08457 [hep-ex].

[34] Armen Tumasyan et al. "Combined searches for the production of supersymmetric top quark partners in proton–proton collisions at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J. C* 81.11 (2021), p. 970. DOI: 10.1140/epjc/s10052-021-09721-5. arXiv: 2107.10892 [hep-ex].

[35] Albert M Sirunyan et al. "Search for top squark pair production using dilepton final states in pp collision data collected at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J. C* 81.1 (2021), p. 3. DOI: 10.1140/epjc/s10052-020-08701-5. arXiv: 2008.05936 [hep-ex].

[36] Georges Aad et al. "Measurement of the polarisation of single top quarks and antiquarks produced in the t-channel at $\sqrt{s} = 13$ TeV and bounds on the tWb dipole operator from the ATLAS experiment". In: *JHEP* 11 (2022), p. 040. DOI: 10.1007/JHEP11(2022)040. arXiv: 2202.11382 [hep-ex].

[37] Morad Aaboud et al. "Search for four-top-quark production in the single-lepton and opposite-sign dilepton final states in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Phys. Rev. D* 99.5 (2019), p. 052009. DOI: 10.1103/PhysRevD.99.052009. arXiv: 1811.02305 [hep-ex].

[38] Kyeongpil Lee et al. "Probing effective field theory operators in the associated production of top quarks with a Z boson in multilepton final states at $\sqrt{s} = 13$ TeV". In: *JHEP* 12 (2021), p. 083. DOI: 10.1007/JHEP12(2021)083. arXiv: 2107.13896 [hep-ex].

[39] The Cms Collaboration et al. "Search for new physics in top quark production with additional leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV using effective field theory". In: *JHEP* 03 (2021), p. 095. DOI: 10.1007/JHEP03(2021)095. arXiv: 2012.04120 [hep-ex].

[40] "Measurement of the polarisation of W bosons produced in top-quark decays using dilepton events at $\sqrt{s} = 13$ TeV with the ATLAS experiment". In: (Sept. 2022). arXiv: 2209.14903 [hep-ex].

[41] Timo Antero Aaltonen et al. "Combined Forward-Backward Asymmetry Measurements in Top-Antitop Quark Production at the Tevatron". In: *Phys. Rev. Lett.* 120.4 (2018), p. 042001. DOI: 10.1103/PhysRevLett.120.042001. arXiv: 1709.04894 [hep-ex].

[42] Morad Aaboud et al. "Combination of inclusive and differential $t\bar{t}$ charge asymmetry measurements using ATLAS and CMS data at $\sqrt{s} = 7$ and 8 TeV". In: *JHEP* 04 (2018), p. 033. DOI: 10.1007/JHEP04(2018)033. arXiv: 1709.05327 [hep-ex].

[43] M. Beneke et al. "Hadronic top-quark pair production with NNLL threshold resummation". In: *Nucl. Phys. B* 855 (2012), pp. 695–741. DOI: 10.1016/j.nuclphysb.2011.10.021. arXiv: 1109.1536 [hep-ph].

[44] Matteo Cacciari et al. "Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation". In: *Phys. Lett. B* 710 (2012), pp. 612–622. DOI: 10.1016/j.physletb.2012.03.013. arXiv: 1111.5869 [hep-ph].

[45] Peter Bärnreuther, Michal Czakon, and Alexander Mitov. "Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to $q\bar{q} \to t\bar{t} + X$". In: *Phys. Rev. Lett.* 109 (2012), p. 132001. DOI: 10.1103/PhysRevLett.109.132001. arXiv: 1204.5201 [hep-ph].

[46] Michal Czakon and Alexander Mitov. "NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels". In: *JHEP* 12 (2012), p. 054. DOI: 10.1007/JHEP12(2012)054. arXiv: 1207.0236 [hep-ph].

[47] Michal Czakon and Alexander Mitov. "NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction". In: *JHEP* 01 (2013), p. 080. DOI: 10.1007/JHEP01(2013)080. arXiv: 1210.6832 [hep-ph].

[48] Michał Czakon, Paul Fiedler, and Alexander Mitov. "Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$". In: *Phys. Rev. Lett.* 110 (2013), p. 252004. DOI: 10.1103/PhysRevLett.110.252004. arXiv: 1303.6254 [hep-ph].

[49] Michal Czakon and Alexander Mitov. "Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders". In: *Comput. Phys. Commun.* 185 (2014), p. 2930. DOI: 10.1016/j.cpc.2014.06.021. arXiv: 1112.5675 [hep-ph].

[50] Michiel Botje et al. "The PDF4LHC Working Group Interim Recommendations". In: (Jan. 2011). arXiv: 1101.0538 [hep-ph].

[51] A. D. Martin et al. "Uncertainties on alpha(S) in global PDF analyses and implications for predicted hadronic cross sections". In: *Eur. Phys. J. C* 64 (2009), pp. 653–680. DOI: 10.1140/epjc/s10052-009-1164-2. arXiv: 0905.3531 [hep-ph].

[52] A. D. Martin et al. "Parton distributions for the LHC". In: *Eur. Phys. J. C* 63 (2009), pp. 189–285. DOI: 10.1140/epjc/s10052-009-1072-5. arXiv: 0901.0002 [hep-ph].

[53] Hung-Liang Lai et al. "New parton distributions for collider physics". In: *Phys. Rev. D* 82 (2010), p. 074024. DOI: 10.1103/PhysRevD.82.074024. arXiv: 1007.2241 [hep-ph].

[54] Jun Gao et al. "CT10 next-to-next-to-leading order global analysis of QCD". In: *Phys. Rev. D* 89.3 (2014), p. 033009. DOI: 10.1103/PhysRevD.89.033009. arXiv: 1302.6246 [hep-ph].

[55] Richard D. Ball et al. "Parton distributions with LHC data". In: *Nucl. Phys. B* 867 (2013), pp. 244–289. DOI: 10.1016/j.nuclphysb.2012.10.003. arXiv: 1207.1303 [hep-ph].

[56] Victor Mukhamedovich Abazov et al. "Measurement of the $t\bar{t}$ production cross section using dilepton events in $p\bar{p}$ collisions". In: *Phys. Lett. B* 704 (2011), pp. 403–410. DOI: 10.1016/j.physletb.2011.09.046. arXiv: 1105.5384 [hep-ex].

[57]   V. M. Abazov et al. "Measurement of the top quark pair production cross section in the lepton + jets channel in proton-antiproton collisions at $\sqrt{s} = 1.96$ TeV". In: *Phys. Rev. D* 84 (1 July 2011), p. 012008. DOI: `10.1103/PhysRevD.84.012008`. URL: `https://link.aps.org/doi/10.1103/PhysRevD.84.012008`.

[58]   T. Aaltonen et al. "Measurement of the Ratio $\sigma_{t\bar{t}}/\sigma_{Z/\gamma^* \to ll}$ and Precise Extraction of the $t\bar{t}$ Cross Section". In: *Phys. Rev. Lett.* 105 (1 June 2010), p. 012001. DOI: `10.1103/PhysRevLett.105.012001`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.105.012001`.

[59]   "Combination of inclusive top-quark pair production cross-section measurements using AT-LAS and CMS data at $\sqrt{s} = 7$ and 8 TeV". In: (May 2022). arXiv: `2205.13830 [hep-ex]`.

[60]   "Measurement of the $t\bar{t}$ production cross-section in pp collisions at $\sqrt{s} = 5.02$ TeV with the ATLAS detector". In: (July 2022). arXiv: `2207.01354 [hep-ex]`.

[61]   Armen Tumasyan et al. "Measurement of the inclusive $t\bar{t}$ production cross section in proton-proton collisions at $\sqrt{s} = 5.02$ TeV". In: *JHEP* 04 (2022), p. 144. DOI: `10.1007/JHEP04(2022)144`. arXiv: `2112.09114 [hep-ex]`.

[62]   Georges Aad et al. "Measurement of the $t\bar{t}$ production cross-section in the lepton+jets channel at $\sqrt{s} = 13$ TeV with the ATLAS experiment". In: *Phys. Lett. B* 810 (2020), p. 135797. DOI: `10.1016/j.physletb.2020.135797`. arXiv: `2006.13076 [hep-ex]`.

[63]   Albert M Sirunyan et al. "Measurement of the $t\bar{t}$ production cross section, the top quark mass, and the strong coupling constant using dilepton events in pp collisions at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J. C* 79.5 (2019), p. 368. DOI: `10.1140/epjc/s10052-019-6863-8`. arXiv: `1812.10505 [hep-ex]`.

[64]   *First measurement of the top quark pair production cross section in proton-proton collisions at $\sqrt{s} = 13.6$ TeV*. Tech. rep. Geneva: CERN, 2022. URL: `https://cds.cern.ch/record/2834110`.

[65]   *Measurement of the $t\bar{t}$ cross-section and $t\bar{t}/Z$ cross-section ratio using LHC Run 3 pp collision data at a centre-of-mass energy of $\sqrt{s} = 13.6$ TeV*. Tech. rep. All figures including auxiliary figures are available at `https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2022-070`. Geneva: CERN, 2022. URL: `https://cds.cern.ch/record/2842916`.

[66]   "Measurement of lepton kinematic distributions in the $e\mu$ decay channel of $t\bar{t}$ events observed in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: (2022).

[67]   *Measurement of differential cross sections for the production of top quark pairs and of additional jets in pp collisions at $\sqrt{s} = 13$ TeV*. Tech. rep. Geneva: CERN, 2022. URL: `https://cds.cern.ch/record/2803771`.

[68] John Campbell, Tobias Neumann, and Zack Sullivan. "Single-top-quark production in the *t*-channel at NNLO". In: *JHEP* 02 (2021), p. 040. DOI: 10.1007/JHEP02(2021)040. arXiv: 2012.01574 [hep-ph].

[69] Richard D. Ball et al. "The PDF4LHC21 combination of global PDF fits for the LHC Run III". In: *J. Phys. G* 49.8 (2022), p. 080501. DOI: 10.1088/1361-6471/ac7216. arXiv: 2203.05506 [hep-ph].

[70] V. M. Abazov et al. "Observation of Single Top Quark Production". In: *Phys. Rev. Lett.* 103 (2009), p. 092001. DOI: 10.1103/PhysRevLett.103.092001. arXiv: 0903.0850 [hep-ex].

[71] T. Aaltonen et al. "First Observation of Electroweak Single Top Quark Production". In: *Phys. Rev. Lett.* 103 (2009), p. 092002. DOI: 10.1103/PhysRevLett.103.092002. arXiv: 0903.0885 [hep-ex].

[72] Serguei Chatrchyan et al. "Measurement of the Single-Top-Quark *t*-Channel Cross Section in *pp* Collisions at $\sqrt{s} = 7$ TeV". In: *JHEP* 12 (2012), p. 035. DOI: 10.1007/JHEP12(2012)035. arXiv: 1209.4533 [hep-ex].

[73] Georges Aad et al. "Comprehensive measurements of *t*-channel single top-quark production cross sections at $\sqrt{s} = 7$ TeV with the ATLAS detector". In: *Phys. Rev. D* 90.11 (2014), p. 112006. DOI: 10.1103/PhysRevD.90.112006. arXiv: 1406.7844 [hep-ex].

[74] Vardan Khachatryan et al. "Measurement of the t-channel single-top-quark production cross section and of the | V_{tb} | CKM matrix element in pp collisions at $\sqrt{s}= 8$ TeV". In: *JHEP* 06 (2014), p. 090. DOI: 10.1007/JHEP06(2014)090. arXiv: 1403.7366 [hep-ex].

[75] Morad Aaboud et al. "Measurement of the inclusive cross-sections of single top-quark and top-antiquark *t*-channel production in *pp* collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *JHEP* 04 (2017), p. 086. DOI: 10.1007/JHEP04(2017)086. arXiv: 1609.03920 [hep-ex].

[76] Morad Aaboud et al. "Fiducial, total and differential cross-section measurements of *t*-channel single top-quark production in *pp* collisions at 8 TeV using data collected by the ATLAS detector". In: *Eur. Phys. J. C* 77.8 (2017), p. 531. DOI: 10.1140/epjc/s10052-017-5061-9. arXiv: 1702.02859 [hep-ex].

[77] Albert M Sirunyan et al. "Measurement of the single top quark and antiquark production cross sections in the *t* channel and their ratio in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Lett. B* 800 (2020), p. 135042. DOI: 10.1016/j.physletb.2019.135042. arXiv: 1812.10514 [hep-ex].

[78] Vardan Khachatryan et al. "Measurement of Top Quark Polarisation in T-Channel Single Top Quark Production". In: *JHEP* 04 (2016), p. 073. DOI: 10.1007/JHEP04(2016)073. arXiv: 1511.02138 [hep-ex].

[79]   Morad Aaboud et al. "Probing the W tb vertex structure in t-channel single-top-quark production and decay in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector". In: *JHEP* 04 (2017), p. 124. DOI: `10.1007/JHEP04(2017)124`. arXiv: `1702.08309` [`hep-ex`].

[80]   Albert M Sirunyan et al. "Measurement of CKM matrix elements in single top quark *t*-channel production in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Lett. B* 808 (2020), p. 135609. DOI: `10.1016/j.physletb.2020.135609`. arXiv: `2004.12181` [`hep-ex`].

[81]   Nikolaos Kidonakis. "Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-". In: *Phys. Rev. D* 82 (2010), p. 054018. DOI: `10.1103/PhysRevD.82.054018`. arXiv: `1005.4451` [`hep-ph`].

[82]   Nikolaos Kidonakis. "Top Quark Production". In: *Helmholtz International Summer School on Physics of Heavy Quarks and Hadrons*. 2014, pp. 139–168. DOI: `10.3204/DESY-PROC-2013-03/Kidonakis`. arXiv: `1311.0283` [`hep-ph`].

[83]   Nikolaos Kidonakis and Nodoka Yamanaka. "Higher-order corrections for *tW* production at high-energy hadron colliders". In: *JHEP* 05 (2021), p. 278. DOI: `10.1007/JHEP05(2021)278`. arXiv: `2102.11300` [`hep-ph`].

[84]   Georges Aad et al. "Evidence for the associated production of a W boson and a top quark in ATLAS at $\sqrt{s} = 7 TeV$". In: *Phys. Lett. B* 716 (2012), p. 142. DOI: `10.1016/j.physletb.2012.08.011`. arXiv: `1205.5764` [`hep-ex`].

[85]   Serguei Chatrchyan et al. "Evidence for associated production of a single top quark and W boson in p-p collisions at $\sqrt{s} = 7 TeV$". In: *Phys. Rev. Lett.* 110 (2013), p. 022003. DOI: `10.1103/PhysRevLett.110.022003`. arXiv: `1209.3489` [`hep-ex`].

[86]   Serguei Chatrchyan et al. "Observation of the associated production of a single top quark and a W boson in p-p collisions at $\sqrt{s} = 8 TeV$". In: *Phys. Rev. Lett.* 112 (2014), p. 231802. DOI: `10.1103/PhysRevLett.112.231802`. arXiv: `1401.2942` [`hep-ex`].

[87]   Georges Aad et al. "Measurement of the production cross-section of a single top quark in association with a W boson at 8 TeV with the ATLAS experiment". In: *JHEP* 01 (2016), p. 064. DOI: `10.1007/JHEP01(2016)064`. arXiv: `1510.03752` [`hep-ex`].

[88]   Timo Antero Aaltonen et al. "Observation of s-channel production of single top quarks at the Tevatron". In: *Phys. Rev. Lett.* 112 (2014), p. 231803. DOI: `10.1103/PhysRevLett.112.231803`. arXiv: `1402.5126` [`hep-ex`].

[89]   Vardan Khachatryan et al. "Search for s channel single top quark production in pp collisions at $\sqrt{s} = 7$ and 8 TeV". In: *JHEP* 09 (2016), p. 027. DOI: `10.1007/JHEP09(2016)027`. arXiv: `1603.02555` [`hep-ex`].

[90]   Georges Aad et al. "Search for *s*-channel single top-quark production in proton–proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector". In: *Phys. Lett. B* 740 (2015), pp. 118–136. DOI: `10.1016/j.physletb.2014.11.042`. arXiv: `1410.0647` [`hep-ex`].

[91] Georges Aad et al. "Evidence for single top-quark production in the *s*-channel in proton-proton collisions at $\sqrt{s}$ =8 TeV with the ATLAS detector using the Matrix Element Method". In: *Phys. Lett. B* 756 (2016), pp. 228–246. DOI: `10.1016/j.physletb.2016.03.017`. arXiv: `1511.05980` `[hep-ex]`.

[92] "Measurement of single top-quark production in the s-channel in proton−proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: (Sept. 2022). arXiv: `2209.08990` `[hep-ex]`.

[93] Georges Aad et al. "Measurement of the $t\bar{t}t\bar{t}$ production cross section in *pp* collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *JHEP* 11 (2021), p. 118. DOI: `10.1007/JHEP11(2021)118`. arXiv: `2106.11683` `[hep-ex]`.

[94] Albert M Sirunyan et al. "Search for production of four top quarks in final states with same-sign or multiple leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J. C* 80.2 (2020), p. 75. DOI: `10.1140/epjc/s10052-019-7593-7`. arXiv: `1908.06463` `[hep-ex]`.

[95] Tomáš Ježo et al. "An NLO+PS generator for $t\bar{t}$ and Wt production and decay including non-resonant and interference effects". In: *Eur. Phys. J. C* 76.12 (2016), p. 691. DOI: `10.1140/epjc/s10052-016-4538-2`. arXiv: `1607.04538` `[hep-ph]`.

[96] Georges Aad et al. "Search for dark matter produced in association with a single top quark in $\sqrt{s} = 13$ TeV *pp* collisions with the ATLAS detector". In: *Eur. Phys. J. C* 81 (2021), p. 860. DOI: `10.1140/epjc/s10052-021-09566-y`. arXiv: `2011.09308` `[hep-ex]`.

[97] Morad Aaboud et al. "Probing the quantum interference between singly and doubly resonant top-quark production in *pp* collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Phys. Rev. Lett.* 121.15 (2018), p. 152002. DOI: `10.1103/PhysRevLett.121.152002`. arXiv: `1806.04667` `[hep-ex]`.

[98] *Studies of $t\bar{t}/tW$ interference effects in $b\bar{b}\ell^+\ell^{-\prime}\nu\bar{\nu}'$ final states with Powheg and MG5_aMC@NLO setups.* Tech. rep. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2021-042. Geneva: CERN, 2021. URL: `https://cds.cern.ch/record/2792254`.

[99] CERN Council. *CERN Council Gives Go-ahead for Large Hadron Collider.* Dec. 2016. URL: `https://press.cern/news/press-release/cern/cern-council-gives-go-ahead-large-hadron-collider`.

[100] Ewa Lopienska. "The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022". In: (2022). General Photo. URL: `https://cds.cern.ch/record/2800984`.

[101] Jean-Luc Caron. "Magnetic field induced by the LHC dipole's superconducting coils.. Champ magnetique cree par les bobines superconductrices des aimants dipolaires du LHC." AC Collection. Legacy of AC. Pictures from 1992 to 2002. 1998. URL: `https://cds.cern.ch/record/841511`.

[102] Simon White. "Luminosity Scans at the LHC. Luminosity Scans at LHC". In: (2011). URL: `https://cds.cern.ch/record/1357865`.

[103] P. Grafström and W. Kozanecki. "Luminosity determination at proton colliders". In: *Prog. Part. Nucl. Phys.* 81 (2015), pp. 97–148. DOI: `10.1016/j.ppnp.2014.11.002`.

[104] CMS Collaboration and Thomas Mc Cauley. "Collisions recorded by the CMS detector on 14 Oct 2016 during the high pile-up fill". CMS Collection. 2016. URL: `https://cds.cern.ch/record/2231915`.

[105] S. Chatrchyan et al. "The CMS Experiment at the CERN LHC". In: *JINST* 3 (2008), S08004. DOI: `10.1088/1748-0221/3/08/S08004`.

[106] Tai Sakuma. "Cutaway diagrams of CMS detector". In: (2019). URL: `https://cds.cern.ch/record/2665537`.

[107] David Barney. "CMS Detector Slice". CMS Collection. 2016. URL: `https://cds.cern.ch/record/2120661`.

[108] William R Leo. *Techniques for nuclear and particle physics experiments: a how-to approach; 2nd ed.* Berlin: Springer, 1994. DOI: `10.1007/978-3-642-57920-2`. URL: `https://cds.cern.ch/record/302344`.

[109] G. Abbiendi et al. "Study of the effects of radiation on the CMS Drift Tubes Muon Detector for the HL-LHC". In: *JINST* 14.12 (2019). Ed. by Marzio Nessi, p. C12010. DOI: `10.1088/1748-0221/14/12/C12010`. arXiv: `1912.06178 [physics.ins-det]`.

[110] S Chatrchyan et al. "Performance of the CMS Drift Tube Chambers with Cosmic Rays". In: *JINST* 5 (2010), T03015. DOI: `10.1088/1748-0221/5/03/T03015`. arXiv: `0911.4855 [physics.ins-det]`.

[111] K. Bos et al. *LHC computing Grid: Technical Design Report. Version 1.06 (20 Jun 2005)*. Technical design report. LCG. Geneva: CERN, 2005. URL: `https://cds.cern.ch/record/840543`.

[112] I Bird et al. *Update of the Computing Models of the WLCG and the LHC Experiments*. Tech. rep. 2014. URL: `https://cds.cern.ch/record/1695401`.

[113] A. M. Sirunyan et al. "Particle-flow reconstruction and global event description with the CMS detector". In: *JINST* 12.10 (2017), P10003. DOI: `10.1088/1748-0221/12/10/P10003`. arXiv: `1706.04965 [physics.ins-det]`.

[114] Wolfgang Adam et al. *Track Reconstruction in the CMS tracker*. Tech. rep. Geneva: CERN, 2006. URL: https://cds.cern.ch/record/934067.

[115] Susanna Cucciarelli et al. *Track reconstruction, primary vertex finding and seed generation with the Pixel Detector*. Tech. rep. Geneva: CERN, 2006. URL: https://cds.cern.ch/record/927384.

[116] Serguei Chatrchyan et al. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *JINST* 9.10 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 [physics.ins-det].

[117] "Performance of the CMS Drift Tubes at the end of LHC Run 2". In: (2019). URL: https://cds.cern.ch/record/2673608.

[118] A. M. Sirunyan et al. "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *JINST* 13.06 (2018), P06015. DOI: 10.1088/1748-0221/13/06/P06015. arXiv: 1804.04528 [physics.ins-det].

[119] Albert M Sirunyan et al. "Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC". In: *JINST* 16.05 (2021), P05014. DOI: 10.1088/1748-0221/16/05/P05014. arXiv: 2012.06888 [hep-ex].

[120] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti-k_t jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[121] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "FastJet User Manual". In: *Eur. Phys. J. C* 72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].

[122] Emil Bols et al. "Jet Flavour Classification Using DeepJet". In: *JINST* 15.12 (2020), P12012. DOI: 10.1088/1748-0221/15/12/P12012. arXiv: 2008.10519 [hep-ex].

[123] CMS Collaboration. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: *JINST 13 (2018) P05011* (2017). DOI: 10.1088/1748-0221/13/05/P05011. arXiv: 1712.07158.

[124] Albert M Sirunyan et al. "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector". In: *JINST* 14.07 (2019), P07004. DOI: 10.1088/1748-0221/14/07/P07004. arXiv: 1903.06078 [hep-ex].

[125] O. Aberle et al. *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2020. DOI: 10.23731/CYRM-2020-0010. URL: https://cds.cern.ch/record/2749422.

[126] Corinne Pralavorio. *Ground-breaking ceremony for the High-Luminosity LHC*. https://home.cern/news/news/accelerators/ground-breaking-ceremony-high-luminosity-lhc. [Consulted on 2022-12-21].

[127] HL-LHC Project. *LHC - HL-LHC Plan.* `https://hilumilhc.web.cern.ch/sites/default/files/HL-LHC_Janvier2022.pdf`. [Consulted on 2022-12-21].

[128] CMS Collaboration. *The Phase-2 Upgrade of the CMS Beam Radiation Instrumentation and Luminosity Detectors.* Tech. rep. This is the final version, approved by the LHCC. Geneva: CERN, 2021. URL: `https://cds.cern.ch/record/2759074`.

[129] *The Phase-2 Upgrade of the CMS Tracker.* Tech. rep. Geneva: CERN, 2017. DOI: `10.17181/CERN.QZ28.FLHW`. URL: `https://cds.cern.ch/record/2272264`.

[130] *The Phase-2 Upgrade of the CMS Barrel Calorimeters.* Tech. rep. This is the final version, approved by the LHCC. Geneva: CERN, 2017. URL: `https://cds.cern.ch/record/2283187`.

[131] *The Phase-2 Upgrade of the CMS Endcap Calorimeter.* Tech. rep. Geneva: CERN, 2017. DOI: `10.17181/CERN.IV8M.1JY2`. URL: `https://cds.cern.ch/record/2293646`.

[132] *The Phase-2 Upgrade of the CMS Muon Detectors.* Tech. rep. This is the final version, approved by the LHCC. Geneva: CERN, 2017. URL: `https://cds.cern.ch/record/2283189`.

[133] Collaboration CMS. *A MIP Timing Detector for the CMS Phase-2 Upgrade.* Tech. rep. Geneva: CERN, 2019. URL: `https://cds.cern.ch/record/2667167`.

[134] *The Phase-2 Upgrade of the CMS Level-1 Trigger.* Tech. rep. Final version. Geneva: CERN, 2020. URL: `https://cds.cern.ch/record/2714892`.

[135] CMS Collaboration. *The Phase-2 Upgrade of the CMS Data Acquisition and High Level Trigger.* Tech. rep. This is the final version of the document, approved by the LHCC. Geneva: CERN, 2021. URL: `https://cds.cern.ch/record/2759072`.

[136] The CMS Muon System. *CMS DT Longevity Studies - new HGCal geometry.* `https://twiki.cern.ch/twiki/bin/view/CMSPublic/DTLongevitynewHGCal`. [Consulted on 2022-12-22].

[137] CERN CMS Group. *CMG Tools.* `https://github.com/CERN-PH-CMG/cmgtools-lite/tree/104X_dev`. 2022.

[138] Rene Brun et al. *root-project/root: v6.18/02.* Aug. 2019. DOI: `10.5281/zenodo.3895860`. URL: `https://doi.org/10.5281/zenodo.3895860`.

[139] Andreas Hoecker et al. "TMVA: Toolkit for Multivariate Data Analysis". In: *PoS* ACAT (2007), p. 040. arXiv: `physics/0703039`.

[140] Lorenzo Moneta et al. 2010. DOI: `10.48550/ARXIV.1009.1003`. URL: `https://arxiv.org/abs/1009.1003`.

[141] CMS Collaboration. *CMS Higgs Combination toolkit.* `https://github.com/cms-analysis/HiggsAnalysis-CombinedLimit`. 2020.

[142] F. James and M. Roos. "Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations". In: *Comput. Phys. Commun.* 10 (1975), pp. 343–367. DOI: `10.1016/0010-4655(75)90039-9`.

[143] "Analytical Method (AM) for DT Trigger Primitive Generation in Phase 2". In: (2020). URL: `https://cds.cern.ch/record/2747124`.

[144] Nicola Pozzobon, Fabio Montecassiano, and Pierluigi Zotto. "Design of a Compact Hough Transform for a New L1 Trigger Primitives Generator for the Upgrade of the CMS Drift Tubes Muon Detector at the HL-LHC". In: *IEEE Transactions on Nuclear Science* 64.6 (2017), pp. 1474–1479. DOI: `10.1109/TNS.2017.2691358`.

[145] A Tapper and Darin Acosta. *CMS Technical Design Report for the Level-1 Trigger Upgrade*. Tech. rep. Additional contacts: Jeffrey Spalding, Fermilab, Jeffrey.Spalding@cern.ch Didier Contardo, Universite Claude Bernard-Lyon I, didier.claude.contardo@cern.ch. 2013. URL: `https://cds.cern.ch/record/1556311`.

[146] P. V. C. Hough. "Machine Analysis of Bubble Chamber Pictures". In: *Conf. Proc. C* 590914 (1959). Ed. by L. Kowarski, pp. 554–558.

[147] Paul V. C. Hough. US3069654A. URL: `https://image-ppubs.uspto.gov/dirsearch-public/print/downloadPdf/3069654`.

[148] O Callot and S Hansmann-Menzemer. *The Forward Tracking: Algorithm and Performance Studies*. Tech. rep. Geneva: CERN, 2007. URL: `https://cds.cern.ch/record/1033584`.

[149] Xin Lin and Kazunori Otobe. "Hough transform algorithm for real-time pattern recognition using an artificial retina camera". In: *Opt. Express* 8.9 (Apr. 2001), pp. 503–508. DOI: `10.1364/OE.8.000503`. URL: `https://opg.optica.org/oe/abstract.cfm?URI=oe-8-9-503`.

[150] Socret Lee. *Lines Detection with Hough Transform*. `https://miro.medium.com/max/1400/1*Cr73Mte5NNgO16D4moKDQg.png`. [Consulted on 2023-01-04].

[151] The CMS Muon System. *CMS DT Longevity Studies*. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/DTLongevity190218`. [Consulted on 2022-12-22].

[152] Carlos Francisco Erice Cid. "Study of beyond the Standard Model processes in leptonic final states with the CMS detector at the Run II of the LHC". Presented 23 Sep 2021. 2021. URL: `http://cds.cern.ch/record/2783303`.

[153] Paolo Nason. "A New method for combining NLO QCD with shower Monte Carlo algorithms". In: *JHEP* 11 (2004), p. 040. DOI: `10.1088/1126-6708/2004/11/040`. arXiv: `hep-ph/0409146`.

[154] Stefano Frixione, Paolo Nason, and Carlo Oleari. "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method". In: *JHEP* 11 (2007), p. 070. DOI: `10.1088/1126-6708/2007/11/070`. arXiv: `0709.2092 [hep-ph]`.

[155]   Simone Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". In: *JHEP* 06 (2010), p. 043. DOI: `10.1007/JHEP06(2010)043`. arXiv: `1002.2581 [hep-ph]`.

[156]   J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *JHEP* 07 (2014), p. 079. DOI: `10.1007/JHEP07(2014)079`. arXiv: `1405.0301 [hep-ph]`.

[157]   Enrico Bothmann et al. "Event Generation with Sherpa 2.2". In: *SciPost Phys.* 7.3 (2019), p. 034. DOI: `10.21468/SciPostPhys.7.3.034`. arXiv: `1905.09127 [hep-ph]`.

[158]   M. Bahr et al. "Herwig++ Physics and Manual". In: *Eur. Phys. J. C* 58 (2008), pp. 639–707. DOI: `10.1140/epjc/s10052-008-0798-9`. arXiv: `0803.0883 [hep-ph]`.

[159]   Johannes Bellm et al. "Herwig 7.0/Herwig++ 3.0 release note". In: *Eur. Phys. J. C* 76.4 (2016), p. 196. DOI: `10.1140/epjc/s10052-016-4018-8`. arXiv: `1512.01178 [hep-ph]`.

[160]   S. Agostinelli et al. "Geant4—a simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: `https://doi.org/10.1016/S0168-9002(03)01368-8`. URL: `https://www.sciencedirect.com/science/article/pii/S0168900203013688`.

[161]   J. Allison et al. "Recent developments in Geant4". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 835 (2016), pp. 186–225. ISSN: 0168-9002. DOI: `https://doi.org/10.1016/j.nima.2016.06.125`. URL: `https://www.sciencedirect.com/science/article/pii/S0168900216306957`.

[162]   J. Allison et al. "Geant4 developments and applications". In: *IEEE Transactions on Nuclear Science* 53.1 (2006), pp. 270–278. DOI: `10.1109/TNS.2006.869826`.

[163]   Christian Bierlich et al. "Robust Independent Validation of Experiment and Theory: Rivet version 3". In: *SciPost Phys.* 8 (2020), p. 026. DOI: `10.21468/SciPostPhys.8.2.026`. arXiv: `1912.05451 [hep-ph]`.

[164]   CMS Collaboration. *Object definitions for top quark analyses at the particle level.* Tech. rep. CMS-NOTE-2017-004. CERN-CMS-NOTE-2017-004. Geneva: CERN, June 2017. URL: `https://cds.cern.ch/record/2267573`.

[165]   Richard D. Ball et al. "Parton distributions from high-precision collider data". In: *Eur. Phys. J. C* 77.10 (2017), p. 663. DOI: `10.1140/epjc/s10052-017-5199-5`. arXiv: `1706.00428 [hep-ph]`.

[166]   Albert M Sirunyan et al. "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements". In: *Eur. Phys. J. C* 80.1 (2020), p. 4. DOI: `10.1140/epjc/s10052-019-7499-4`. arXiv: `1903.12179 [hep-ex]`.

[167] Vardan Khachatryan et al. "Event generator tunes obtained from underlying event and multiparton scattering measurements". In: *Eur. Phys. J. C* 76.3 (2016), p. 155. DOI: `10.1140/epjc/s10052-016-3988-x`. arXiv: `1512.00815 [hep-ex]`.

[168] "CMS PYTHIA 8 colour reconnection tunes based on underlying-event data". In: (May 2022). arXiv: `2205.02905 [hep-ex]`.

[169] "Performance of electron and photon reconstruction in Run 2 with the CMS experiment". In: (2020). URL: `https://cds.cern.ch/record/2725004`.

[170] Serguei Chatrchyan et al. "Performance of CMS Muon Reconstruction in *pp* Collision Events at $\sqrt{s} = 7$ TeV". In: *JINST* 7 (2012), P10002. DOI: `10.1088/1748-0221/7/10/P10002`. arXiv: `1206.4071 [physics.ins-det]`.

[171] "Muon reconstruction performance during Run II". In: (2019). URL: `https://cds.cern.ch/record/2682902`.

[172] *Jet algorithms performance in 13 TeV data*. Tech. rep. Geneva: CERN, 2017. URL: `https://cds.cern.ch/record/2256875`.

[173] Olena Karacheban. "Luminosity measurement at CMS". Presented 05 Jan 2017. 2017. URL: `https://cds.cern.ch/record/2294183`.

[174] S van der Meer. *Calibration of the effective beam height in the ISR*. Tech. rep. Geneva: CERN, 1968. URL: `https://cds.cern.ch/record/296752`.

[175] Albert M Sirunyan et al. "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV". In: *JHEP* 07 (2018), p. 161. DOI: `10.1007/JHEP07(2018)161`. arXiv: `1802.02613 [hep-ex]`.

[176] A. Bodek et al. "Extracting Muon Momentum Scale Corrections for Hadron Collider Experiments". In: *Eur. Phys. J. C* 72 (2012), p. 2194. DOI: `10.1140/epjc/s10052-012-2194-8`. arXiv: `1208.3710 [hep-ex]`.

[177] Vardan Khachatryan et al. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV". In: *JINST* 12.02 (2017), P02014. DOI: `10.1088/1748-0221/12/02/P02014`. arXiv: `1607.03663 [hep-ex]`.

[178] A. M. Sirunyan et al. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: *JINST* 13.05 (2018), P05011. DOI: `10.1088/1748-0221/13/05/P05011`. arXiv: `1712.07158 [physics.ins-det]`.

[179] Albert M Sirunyan et al. "Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS". In: *Eur. Phys. J. C* 81.9 (2021), p. 800. DOI: `10.1140/epjc/s10052-021-09538-2`. arXiv: `2104.01927 [hep-ex]`.

[180] *CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV*. Tech. rep. Geneva: CERN, 2018. URL: `https://cds.cern.ch/record/2621960`.

[181] *CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV*. Tech. rep. Geneva: CERN, 2019. URL: https://cds.cern.ch/record/2676164.

[182] Peter Skands, Stefano Carrazza, and Juan Rojo. "Tuning PYTHIA 8.1: the Monash 2013 Tune". In: *Eur. Phys. J. C* 74.8 (2014), p. 3024. DOI: 10.1140/epjc/s10052-014-3024-y. arXiv: 1404.5630 [hep-ph].

[183] *Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of* $t\bar{t}$ *at $\sqrt{s} = 8$ and 13 TeV*. Tech. rep. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2235192.

[184] Vardan Khachatryan et al. "Measurement of the top quark mass using proton-proton data at $\sqrt{(s)} = 7$ and 8 TeV". In: *Phys. Rev. D* 93.7 (2016), p. 072004. DOI: 10.1103/PhysRevD.93.072004. arXiv: 1509.04044 [hep-ex].

[185] Vardan Khachatryan et al. "Measurement of differential cross sections for top quark pair production using the lepton+jets final state in proton-proton collisions at 13 TeV". In: *Phys. Rev. D* 95.9 (2017), p. 092001. DOI: 10.1103/PhysRevD.95.092001. arXiv: 1610.04191 [hep-ex].

[186] Vardan Khachatryan et al. "Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV". In: *Eur. Phys. J. C* 75.11 (2015), p. 542. DOI: 10.1140/epjc/s10052-015-3709-x. arXiv: 1505.04480 [hep-ex].

[187] Vardan Khachatryan et al. "Measurement of the $t\bar{t}$ production cross section in the all-jets final state in pp collisions at $\sqrt{s} = 8$ TeV". In: *Eur. Phys. J. C* 76.3 (2016), p. 128. DOI: 10.1140/epjc/s10052-016-3956-5. arXiv: 1509.06076 [hep-ex].

[188] *Measurement of inclusive W and Z boson production cross sections in pp collisions at sqrt(s)=13 TeV*. Tech. rep. Geneva: CERN, 2015. URL: https://cds.cern.ch/record/2093537.

[189] J. S. Conway. "Incorporating nuisance parameters in likelihoods for multisource spectra". In: *Proc. 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding (PHYSTAT 2011): Geneva, Switzerland, January 17–20, 2011*. 2011. DOI: 10.5170/CERN-2011-006.115. arXiv: 1103.0354 [physics.data-an].

[190] ATLAS and CMS Collaborations, and LHC Higgs Combination Group. *Procedure for the LHC Higgs boson search combination in Summer 2011*. Tech. rep. CMS-NOTE-2011-005, ATL-PHYS-PUB-2011-11. 2011. URL: https://cds.cern.ch/record/1379837.

[191] Roger J. Barlow and Christine Beeston. "Fitting using finite Monte Carlo samples". In: *Comput. Phys. Commun.* 77 (1993), p. 219. DOI: 10.1016/0010-4655(93)90005-W.

[192] Yoav Freund and Robert E Schapire. "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting". In: *Journal of Computer and System Sciences* 55.1 (1997), pp. 119–139. ISSN: 0022-0000. DOI: `https://doi.org/10.1006/jcss.1997.1504`. URL: `https://www.sciencedirect.com/science/article/pii/S002200009791504X`.

[193] Leo Breiman et al. *Classification and regression trees*. Chapman and Hall, 1984. ISBN: 0-412-048141-8.

[194] Armen Tumasyan et al. "Measurement of differential $t\bar{t}$ production cross sections in the full kinematic range using lepton+jets events from proton-proton collisions at $\sqrt{s} = 13$TeV". In: *Phys. Rev. D* 104 (2021), p. 092013. DOI: `10.1103/PhysRevD.104.092013`. arXiv: `2108.02803` [`hep-ex`].

[195] ATLAS Collaboration. "Measurement of the cross-section for producing a W boson in association with a single top quark in *pp* collisions at $\sqrt{s} = 13$ TeV with ATLAS". In: *JHEP 01 (2018) 63* (2016). DOI: `10.1007/JHEP01(2018)063`. arXiv: `1612.07231`.

[196] Armen Tumasyan et al. "Observation of tW production in the single-lepton channel in pp collisions at $\sqrt{s} = 13$ TeV". In: *JHEP* 11 (2021), p. 111. DOI: `10.1007/JHEP11(2021)111`. arXiv: `2109.01706` [`hep-ex`].

[197] W. Adam et al. "The CMS Phase-1 Pixel Detector Upgrade". In: *JINST* 16.02 (2021), P02027. DOI: `10.1088/1748-0221/16/02/P02027`. arXiv: `2012.14304` [`physics.ins-det`].

[198] *CMS Luminosity Measurements for the 2016 Data Taking Period*. Tech. rep. Geneva: CERN, 2017. URL: `https://cds.cern.ch/record/2257069`.

[199] NNPDF Collaboration et al. "Parton distributions for the LHC Run II". In: *Journal of High Energy Physics 4* (2014). DOI: `10.1007/JHEP04(2015)040`. arXiv: `1410.8849`.

[200] Christian Hansen. "Analysis of discrete ill–posed problems by means of the L-Curve". In: *SIAm Review Vol.34, No.4, pp. 561-580* (1992).

[201] Stefan Schmitt. "TUnfold: an algorithm for correcting migration effects in high energy physics". In: *JINST* 7 (2012), T10003. DOI: `10.1088/1748-0221/7/10/T10003`. arXiv: `1205.6201` [`physics.data-an`].

[202] Albert M Sirunyan et al. "Measurement of the top quark polarization and $t\bar{t}$ spin correlations using dilepton final states in proton-proton collisions at $\sqrt{s} = 13$TeV". In: *Phys. Rev. D* 100 (2019), p. 072002. DOI: `10.1103/PhysRevD.100.072002`. arXiv: `1907.03729` [`hep-ex`].

[203] ATLAS Collaboration. "Measurement of differential cross-sections of a single top quark produced in association with a W boson at $\sqrt{s} = 13$ TeV with ATLAS". In: *Eur. Phys. J. C 78 (2018) 186* (2017). DOI: `10.1140/epjc/s10052-018-5649-8`. arXiv: `1712.01602`.

[204] Stefano Frixione et al. "Single-top hadroproduction in association with a W boson". In: *JHEP* 07 (2008), p. 029. DOI: 10.1088/1126-6708/2008/07/029. arXiv: 0805.3067 [hep-ph].

A Study of the new $b\bar{b}\ell^+\nu_{\ell+}\ell^-\nu_{\ell-}$ sample

As mentioned in Sect. 1.4.3 and showed in the main body of the thesis, experimental measurements of the tW process (or of any other that might have tW as background) are done with MC simulations modified, generally by the DR method. However, a new simulation of the final state $b\bar{b}\ell^+\nu_{\ell+}\ell^-\nu_{\ell-}$ has been developed that models some of the WWbb process final states, allowing to replace the $t\bar{t}$ and tW simulation samples by it [95].

This production mode, part of a `POWHEG`'s plugin, is installed in the CMS software for some years. Initial studies of the generated samples have been done to compare the modelling they provided against the separated tW and $t\bar{t}$ samples, and also to assess the status of the sample, due to its novelty. These efforts allowed to correct issues in its generation and prepare the ground for significant comparisons with the current tW and $t\bar{t}$ modelling. In this section particle-level comparisons of the current $b\bar{b}\ell^+\nu_{\ell+}\ell^-\nu_{\ell-}$ sample with respect to the separated tW (with various different generators) and $t\bar{t}$ samples are presented.

A.1 Simulations

Apart from the $b\bar{b}\ell^+\nu_{\ell+}\ell^-\nu_{\ell-}$ sample, individual simulations of $t\bar{t}$ using `POWHEG` +`Pythia 8` at NLO+PS in QCD are used for comparison. Their contributions are added to several tW process samples, with NLO+PS precision in QCD and using for the PS `Pythia 8` 8 in all cases. The ME is modelled with `POWHEG` DR, and DS, as well as with `MadGraph5_aMC@NLO` DR, DR2, DS, and DS with a dynamic scale.

A.2 Objects and event selection

The identification of particle-level objects is the same as in the case of the analysis of the thesis, and is summarised in Table 7.2. Objects are defined following Sect. 3.2.1 guidelines. A fiducial region is built also similarly to the described analysis, by requiring two dressed leptons (an electron and a muon) of opposite sign, one with $p_{\mathrm{T}} > 25$ GeV; in addition their invariant mass must be > 20 GeV. Then, (exactly) two b-tagged jets, cleaned as in the tW measurement with the dressed leptons, are required. This selection allows to focus in the 2j2b phase-space subregion, where one could expect a larger contribution from the interference effects between the $t\bar{t}$ and tW process because of the signature of two on-shell top quarks.

A.3 Comparisons and discussion

The comparisons are done as a function of the leading lepton $p_{\rm T}$ and η, the leading jet $p_{\rm T}$ and η, the $p_{\rm T}(e^{\pm}, \mu^{\mp})$, and

$$m^{\rm minimax} = \min\left(\max(m_{j_1,\ell_1}, m_{j_2,\ell_2}), \max(m_{j_1,\ell_2}, m_{j_2,\ell_1})\right), \tag{A.1}$$

defined as in [97], where j_i and ℓ_i refer to the jets and leptons of each event and $i = 1$ is the leading object in terms of $p_{\rm T}$, while $i = 2$ is the subleading. This variable, for WWbb events with two on-shell tops verifies that $m^{\rm minimax} < \sqrt{m_t^2 - m_W^2} \approx 152$ GeV. Thus, interference effects should be larger above that threshold than below, allowing us to probe the differences in modelling with the separated $t\bar{t}$ and tW.

Figure A.1 show the comparisons in these six distributions between the $b\bar{b}\ell^+\nu_{\ell^+}\ell^-\nu_{\ell^-}$ sample and the separated tW and $t\bar{t}$ simulations, normalised to the fiducial cross section. There are clear shape differences in the modelling of the η of the leading jet and lepton between all the separated $t\bar{t}$ +tW samples and $b\bar{b}\ell^+\nu_{\ell^+}\ell^-\nu_{\ell^-}$. There are also some variations at the end of $p_{\rm T}(e^{\pm}, \mu^{\mp})$, partly expected as shown in [204]. Finally, the $m^{\rm minimax}$ variable shows as expected contrasts between $b\bar{b}\ell^+\nu_{\ell^+}\ell^-\nu_{\ell^-}$ and the separated-processes samples after the mentioned threshold, growing towards the tail of the distributions. The differences are comparable to those observed in [97].
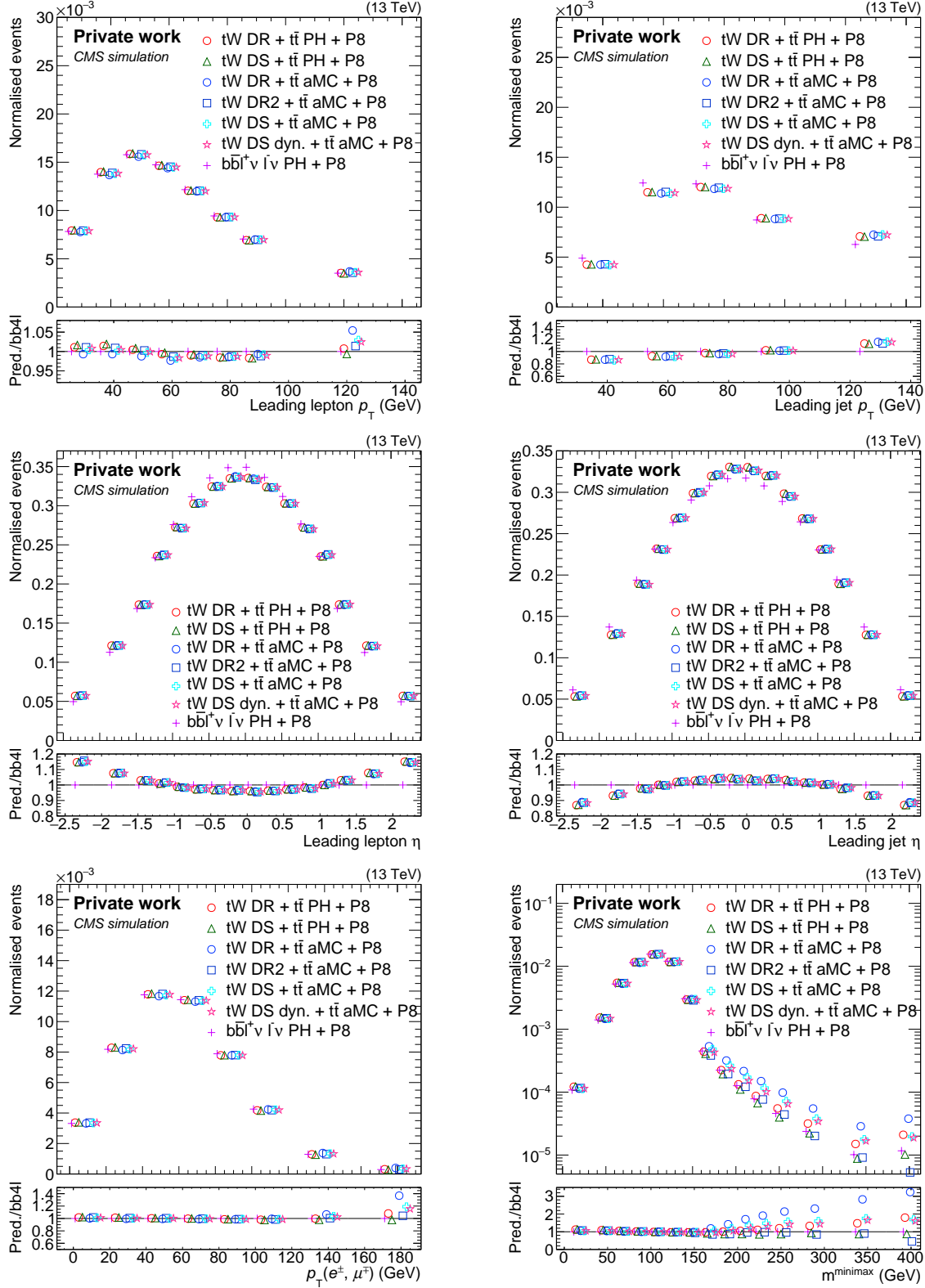
Figure A.1: Comparisons at particle-level between the $b\bar{b}\ell^+\nu_{\ell+}\ell^-\nu_{\ell-}$ sample and various combinations of different tW contributions added up to $t\bar{t}$ ones. They are shown depending on the p_T of the leading lepton (upper left) and leading jet (upper right), η of the leading lepton (middle left) and leading jet (middle right), $p_T(e^\pm, \mu^\mp)$ (lower left), and m^{minimax} (lower right).